Article

Different Viewpoints Image Registration for Remote Sensing Based on Multiple Image Features

Kun Yang [1,2,†], Anning Pan [1,2,†], Yang Yang [1,2,*], Su Zhang [1,2] and Sim Heng Ong [3,4,5]

[1] School of Information Science and Technology, Yunnan Normal University, Kunming 650500, Yunnan, China; kmdcynu@163.com (A.P.); paninglw@163.com (Y.Y.); sorazcn@gmail.com (S.Z.)

[2] The Engineering Research Center of GIS Technology in Western China of Ministry of Education of China, Yunnan Normal University, Kunming 650500, Yunnan, China

[3] NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore 117456, Singapore; eleongsh@nus.edu.sg

[4] Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576, Singapore

[5] Department of Bioengineering, National University of Singapore, Singapore 117576, Singapore

* Correspondence: yyang_ynu@163.com; Tel.: +1-831-445-2528

† These authors contributed equally to this work.

Abstract: Remote sensing image registration with different viewpoints plays an important role in the field of geographic information system. However, when there exists ground relief variations and imaging viewpoint changes, non-rigid distortion occurs thus the registration becomes increasingly challenging. The current methods will suffer from missing true correspondences when non-rigid geometric distortion occurs. To address the problem, we propose a robust remote sensing image registration method based on SIFT feature distance and geometric structure features. At first, the scale-invariant feature transform (SIFT), a partial intensity invariant feature descriptor is used to extract reliable feature point set from sensed and reference image respectively. Secondly, a novel algorithm based on multiple image features which constrains the geometric structure during transformation is used to estimate exact correspondences between point sets. Finally, an accurate alignment is achieved by mapping the sensed image to reference image using thin-plate spline. We evaluated the performances of the proposed method by three sets of remote sensing images obtained from the unmanned aerial vehicle (UAV) and the Google earth, and compared with five state-of-the-art methods where our algorithm solved the non-rigid registration problem of remote sensing image with different viewpoints and showed the best alignments in most cases.

Keywords: remote sensing; image registration; multiple image features; different viewpoints; non-rigid distortion

1. Introduction

In recent years, image registration has become a extremely important technique in a wide range of applications such as computer vision, pattern recognition, environment monitoring, medical image analysis and remote sensing. Image registration refers to the fundamental task in image processing to align two or more images of the same scene (i.e., the reference images and the sensed images), which can be multitemporal (taken at different times), multisource (derived from different sensors) and multiview (obtained from different viewpoints). In this paper, we mainly focus on registering the remote sensing images taken from different viewpoints. As a basic process it becomes a vital yet challenging problem to find more accurate registration algorithms. With an exact registration for remote sensing, it will be well used in many important applications, both at social and scientific levels. These applications include, for example, military automatic target recognition, compiling and analyzing images and data from satellites, assignment of climate changes, environment monitoring, and the management of nature disasters [1].

Existing remote sensing image registration methods can be approximately classified into two categories: area-based methods and feature-based methods. In [2–6], various reviews on image registration methods can be found. We briefly review them here, especially in the application of remote sensing image registration.

The area-based methods can be broadly classified into three types: correlation-like methods, Fourier methods, and mutual information (MI) methods [7]. In correlation-like methods area, an automatic image registration method was proposed by Goncalves et al. [8], which is based on the identification of a thin line through the Hough transform. Based on the singular value decomposition (SVD) and the unified random sample consensus (RANSAC) algorithm, Tong et al. [9] presented a novel subpixel phase correlation method which demonstrated the promising performance and feasibility. However, The correlation-like methods suffer from the high computational complexity and the flatness of the similarity measure in textureless regions. Fourier based method is a popular approach for coarse registration, it is a frequency estimation problem in the frequency domain of the image data essentially. By employing the multiple signal classifier algorithm, Xu et al. [10] studied the application of a subspace-based frequency estimation approach for the Fourier-based image registration problem and achieved a more robust and accurate registration result. The Fourier based methods have less computational complexity, however, considerable spectral contents differences existed in image pairs can cause a drop in performance. MI methods has recently been used as a similarity measure for remote sensing and medical image registration, which possesses the characters of high accuracy and better generality. In order to register multitemporal remote sensing images, Chen et al. [11] demonstrated a new joint histogram estimation algorithm called generalized partial volume estimation (GPVE) for computing mutual information, the method produced a better registration consistency. In addition, an approach based on the implementation of particle swarm optimization (PSO) and mutual information (MI) is proposed to determine more accurate pairs of corresponding points between the images. Nevertheless, the accuracy of the MI-based methods decreases since these methods do not provide a global maximum of the entire search space for the transformation.

Area-based methods usually compare intensity patterns in images via correlation metrics and register entire images or sub-images, while feature-based methods find correspondence between image features such as points, lines, and contours, and establish a correspondence between a number of especially distinct points in images [12]. The feature-based methods are based on pixel intensities instead of local shapes and structures, they perform robustly against noise contamination, rotation and illumination variations and multisensor analysis. In addition, when there is complex distortion between the images to be aligned, the computation complexity or the search space of the area-based methods will increase nonlinearly with the transformation complexity and feature-based methods are preferable. In this work, the remote sensing images captured exist the local no-rigid geometric distortions caused by ground relief variations and imaging viewpoint changes, thus we mainly focus on feature-based methods (i.e., local invariant features) for registration. The method generally consists of four steps [13]: (1) feature descriptors extraction; (2) Feature point sets registration; (3) Transformation-model estimation; (4) Image transformation and resampling.

There are a variety of popular local invariant features which have been proposed in remote sensing image registration, such as scale-invariant feature transform (SIFT) [14–17], speeded-up robust features (SURF) [18] and Harris [19], etc. Recently, the computational efficiency and registration accuracy have been improved by optimizing the feature extraction and adding extra constraint for feature matching based on local invariant features [12][18][20–23]. Generally, a good descriptor should have two main properties which are distinctiveness (different features should have different descriptors) and robustness (a descriptor's stability to a variety of image geometric and photometric transformations). Improving distinctiveness while maintaining robustness is the main concern in the design of local image descriptors [24]. Having a comparison of the performance between different descriptors for affine transformation, scale change, rotation, image blur, jpeg compression, and illumination change, Mikolajczyk et al. [25] demonstrated that scale-invariant feature transform (SIFT) [14] performs the best for most of the tests.

SIFT as a capable feature descriptor of extracting distinctive invariant features from images, it can be applied to perform reliable registration across a substantial range of affine distortion, change in 3-D viewpoint, addition of noise, and change in illumination [15]. Therefore, there are various researches for remote sensing image registration based on the variants of SIFT algorithm and the combination of SIFT and some other algorithms. Li et al. [26] proposed a new criterion named scale-orientation joint restriction criteria in order to overcome the intensity difference between remote sensing image pairs. Sedaghat et al. [27] introduced an automatic registration algorithm by extracting high-quality SIFT features in the uniform distribution of both the scale and image spaces. Furthermore, some variants of SIFT such as PCA-SIFT [28], SAR-SIFT [29], AB-SIFT [30] and SIFT-DRS [31] are proposed in order to make an improvement on SIFT. Furthermore, Goncalves et al. [32] developed a new AIR (Automatic image registration) method based on the combination of image segmentation and SIFT, and the method is complemented by a robust procedure of outlier removal. In [33], a novel coarse-to-fine scheme for automatic image registration based on SIFT and MI is proposed, and their method achieved the outlier removal and also can generally reject most incorrect matches.

However, if the image pairs are obtained under a large difference of camera viewpoints and suffer from local distortion caused by ground relief variations and imaging viewpoint changes ,the correspondences obtained by estimating only SIFT feature can be unreliable.

Feature-based methods are typically formulated as a point set registration problem since point representations are general and easy to extract. In order to achieve a robust point registration, it is crucial to construct putative correspondences based on local invariant feature similarity at first and then estimate the spatial transformation based on geometric structure feature constraint [1][34]. Here, we briefly review some point set registration methods since our method is based on feature point. Fischler et al. proposed the classical Random Sample Consensus (RANSAC) algorithm [35]. Myronenko and Song [36] introduced a probabilistic method, called the Coherent Point Drift (CPD) algorithm, for both rigid and non-rigid point set registration. Moreover, Liu and Yan [37] investigated how to discover common visual pattern discovery via spatially coherent correspondences and recover the correct correspondences. The traditional LDA models which are used to solve the problem of scene classification lack the spatial relationship and linkages between the global and local information, so Zhang et al. [38] defined the Spatially Consistent Topic Model by making full use of the correlation between image classification and annotation. Recently, Ma et al. proposed a robust L2-minimizing estimate (L2E) [39] for non-rigid point set registration, they later proposed a flexible and general algorithm called locally linear transforming (LLT) [40] for both rigid and non-rigid registration on remote sensing images. More recently, a new method named GLMDTPS was proposed by Yang [41]. However, these algorithms consider only the geometric features without combining with meaningful local invariant features, which makes the features of points be less distinctive.

Although many approaches above-mentioned have been proposed for different applications, there still exists the following problems for remote sensing image registration at present. First, most of the methods mainly focus on solving the rigid and affine geometric distortion problems, which means that the non-rigid distortion problems exacerbated by the complexity of space distribution and terrestrial objects are in urgent need to be settled. Second, the methods mainly focused on single feature and have a limited performance when directly applied to remote sensing images. For example, [14–17] achieved image registration through SIFT only. Although some variants of SIFT algorithm in [26–32] are proposed, these algorithms are still relatively limited as they do not consider the geometric structure features. Severe non-rigid distortion can significantly deteriorate the correspondence estimation on the remote sensing image pair. Furthermore, literatures [34–41] only use the geometric structure features between the correspondence and the transformation estimation.

Compared with the current methods, the major differences and advantages of this work include: (i) The Euclidean distance and the shape context [42] descriptor are respectively used as the global and local geometric features, both of which play complementary roles to exploit the geometric structure of the feature point sets. (ii) We combine the SIFT feature, a local invariant feature with geometric structure features to form the multiple image feature, this feature is the base upon which our robust remote sensing image registration method estimates the correspondence. (iii) To constrain the non-rigid transformation which is too arbitrary, the mapping is regularized by a global structure preservation term which further improves the accuracy of the registration.

The rest of this paper is organized as follows: section 2 introduces our method in detail. Section 3 demonstrates the registration performance of our method on various types of remote sensing images with different viewpoints against other approaches, followed by some concluding remarks in section 4. Results show that the proposed method is highly distinctive and robust under non-rigid geometric distortions.

2. Methodology

2.1. Feature Descriptors Extraction

At the first step, given a remote sensing image I_r (i.e., the reference image) and another one with the different viewpoint I_s (i.e., the sensed image), we employ the SIFT detector [14] to extract feature points from the reference and the sensed images respectively. The SIFT algorithm has been introduced by Lowe in 1999 [14] and then improved in 2004 [15] in order to match local features in natural images. We now introduce the SIFT algorithm briefly in this section, it usually contain the following steps:

- Keypoints Detection: At first, keypoints are selected and characterized by their position (x,y), scale σ, and orientation θ, i.e.,
$$P(x, y, \sigma, \theta).$$
By using difference-of-gaussians, the pixel value extrema in both scale and space are found.
- Keypoint localization: in this step, the most stable extrema are converted to keypoints.
- Orientation Assignment: Lowe [15] proposed to compute a local histogram of gradient orientations, weighted by the gradient magnitudes and a Gaussian window. The most dominant direction of the histogram is detected and used as the orientation θ of the keypoint. If the peak is above 80% of the maximum then it is selected as an orientation. We can obtain several keypoints with the same position (x,y) and scale σ but with different orientations θ.
- Descriptors Extraction: the local image gradients are stored to represent each keypoint $P(x, y, \sigma, \theta)$. In order to get translation and scale invariance, a square neighborhood is defined around each point with a size depending on σ and θ, where θ can ensure its rotation invariance. The aforementioned normalized neighborhood consists of 4×4 histograms, in which all of the orientations are summarized. Each histogram exists of 8 bins which represent the gradients in that directions. Then for each keypoint, the $4 \times 4 \times 8$ histogram bins make a feature vector with 128 values. Before computing these histograms, a Gaussian weighting function is applied to reduce the effect of pixels further away from the keypoint. Then the SIFT descriptor is obtained via concatenating and normalizing these histograms for each keypoint.

In this paper, we choose to use a SIFT detector to extract feature points from remote sensing images with different viewpoints, since the distinctiveness is promising, which make it scale and rotation invariant and performance better to a large range of affine distortion and illumination changes. What more, with this outstanding property, we believed that each feature vector with 128 values greatly help feature point sets registration, so a new SIFT feature distance is presented to improve the robustness and accuracy of transformation. More details will be elaborated later.

2.2. Feature Point Sets Registration

After obtained feature point sets $\mathbf{Y}_{M\times D} = \{y_1, ..., y_M\}^T$ and $\mathbf{X}_{N\times D} = \{x_1, ..., x_N\}^T$ from the sensed image and the reference image where D denotes the dimension of the point set, and the corresponding point set be denoted by placing a hat above the letter. we define a warping template $\mathbf{Y}^w$ (the initial $\mathbf{Y}^w = \mathbf{Y}$) to estimate corresponding points $\hat{\mathbf{X}}$ (the points in X). And our new non-rigid point set registration method is based on the iteration of the following two main steps: (i) estimating correspondence between $\mathbf{Y}$ and $\hat{\mathbf{X}}$; (ii) Mapping $\mathbf{Y}^w$ to the correspondent $\mathbf{X}$ until resulting in the maximum point-wise overlap between two point sets. Here we consider point set registration as a probability density estimation problem.

2.2.1. Correspondence Estimation

- **Gaussian Mixture Model(GMM):** The Gaussian distribution is one of the most famous probability statistic models. The GMM probability density function is given by

$$p(\mathbf{x}_n) = \sum_{m=1}^{M+1} P(m)p(\mathbf{x}_n|\theta_m), \tag{1}$$

where $P(m)$ denotes the mixing weights. In the meantime, all GMM components share the equal isotropic covariances and equal membership probabilities. Furthermore, in order to account for noise and outliers, we added an additional uniform distribution $p(x|M+1) = \frac{1}{N}$ to the mixture model . Where ω, $0 \leq \omega \leq 1$, denotes the weight of the uniform distribution and the GMM density function takes the form

$$p(\mathbf{x}_n) = (1-\omega)\sum_{m=1}^{M} P(m)p(\mathbf{x}_n|\boldsymbol{\theta}_m) + \omega\frac{1}{N}, \tag{2}$$

Then we rewrite the posterior probability function of GMM in the form

$$p_{mn} = \frac{\exp(-\frac{1}{2\sigma^2}\|\mathbf{x}_n - \mathbf{y}_m^w\|^2)}{\sum_{j=1}^{M}\exp(-\frac{1}{2\sigma^2}\|\mathbf{x}_n - \mathbf{y}_j^w\|^2) + (2\pi\sigma^2)^{\frac{D}{2}}\frac{\omega}{1-\omega}\frac{M}{N}}, \tag{3}$$

However, consider only the Euclidean distance between $\mathbf{X}$ and $\mathbf{Y}^w$ in (3) may lead to insufficient robustness in some registration scenarios. For example, although the local structure around $\mathbf{x}_a$ and $\mathbf{x}_b$ are probably totally different, if $\|\mathbf{x}_a - \mathbf{y}_m^w\|^2 = \|\mathbf{x}_b - \mathbf{y}_m^w\|^2$, we obtain $p_{ma} = p_{mb}$. Shape context is another geometric feature which is presented for describing the local difference. And to settle this problem, we consider SIFT feature as a local invariant feature to combine with geometric features.

- **Mixture-feature Gaussian Mixture Model:** Firstly, let us define the Gaussian function in the form

$$p(x|\theta) = w\exp\left(-\frac{\|\mathbf{x} - \mathbf{y}^w\|^2}{2\sigma^2}\right). \tag{4}$$

By changing either the constant w or the exponent term $\|\mathbf{x} - \mathbf{y}^w\|^2$, we can obtain a different $p(x|\theta)$. Due to the exponent term $\|\mathbf{x} - \mathbf{y}^w\|^2$ is much easier to insert extra terms. So we aim to implement the adjustment on the exponent term of (3), and define the mixture feature descriptor of point set as

$$\mathcal{F} = \mathbf{G}^{Global} + \mathbf{G}^{Local} + \mathbf{S}, \tag{5}$$

where $\mathbf{G}^{Global}$ is global geometric structure features and $\mathbf{G}^{Local}$ is local geometric structure features, $\mathbf{S}$ denotes local invariant feature distance and it is a $M \times N$ matrix. Though the European distance feature is not robust, it is the basis for making the model effective, is essential and taken as global geometric structure features. The shape context is considered as local geometric structure features because it can supplement the above-mentioned deficiencies, distinguishing the situation of insufficient robustness in some registration scenarios. Local invariant features can enhance the ability to distinguish, SIFT feature is selected in our paper. We will explain each of them in the following.

- **Geometric Structure Features:** In this work, global geometric structure features is used to describe squared Euclidean distances from one point to another, and defined as

$$\mathbf{G}_{mn}^{Global} = \|\mathbf{x}_n - \mathbf{y}_m^w\|^2 \tag{6}$$

Shape context [42] is used as a local geometric structure descriptor. In considering two points, $\mathbf{x}_n$ is in one set and $\mathbf{y}_m$ is in another set; their shape contents are $\mathbf{h}_n(k)$ and $\mathbf{h}_m(k)$, respectively, where h(k) is the value of k-bin for the log-polar histogram. Let c_{mn} represent the matching cost measure of these two points. Then the local geometric structure feature is defined as:

$$\mathbf{G}_{mn}^{Local} = \mathbf{c_{mn}} = \frac{1}{2}\sum_{k=1}^{K}\frac{(\mathbf{h}_n(k) - \mathbf{h}_m(k))^2}{\mathbf{h}_n(k) + \mathbf{h}_m(k)}. \tag{7}$$

The more similar the shape contexts for the two points of $\mathbf{x}_n$ and $\mathbf{y}_m$ are, the more likely for these two point to be corresponding. Thus the combination of the aforementioned global and local geometric structure features difference formed a mixed geometric structure features which is defined as $\mathbf{G} = \mathbf{G}_{mn}^{Global} + \gamma\mathbf{G}_{mn}^{Local}$.

- **SIFT Feature Distance:** SIFT can transform an image into a large collection of local feature vectors and each of which is invariant to image rotation, scaling, translation and partially invariant to affine, 3D projection or illumination changes. In our work, SIFT feature distance is used to enhance the distinctiveness of each point and defined as:

$$\mathbf{S}_{\mathbf{x}_n\mathbf{y}_m^{sift}} = (\mathbf{x_i} - \mathbf{y_j})^2. \tag{8}$$

Where the matrix $\mathbf{S_{xy}}$ describes the SIFT feature distance between point set x and point set y. It is considered as a feature vector with 128 values for each point set in X and Y. Each descriptor vector can be considered as the following form: $\mathbf{y}_j = \{y_1, y_2, ..., y_{128}\}$ and $\mathbf{x}_i = \{x_1, x_2, ..., x_{128}\}$.

Substituting (5) into (4), it takes the form

$$\begin{aligned} p(x|\theta) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\mathbf{G}^{Global} - \gamma\mathbf{G}^{Local} - \xi\mathbf{S}^{sift}\right)^2 \\ &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\|x - \mathbf{y}^w\|^2}{2\sigma^2} - \frac{\gamma}{2}\sum_{k=1}^{K}\frac{(\mathbf{h}_n(k) - \mathbf{h}_m(k))^2}{\mathbf{h}_n(k) + \mathbf{h}_m(k)} - \xi(\mathbf{x_i} - \mathbf{y_j})^2\right)^2. \end{aligned} \tag{9}$$

Where σ^2 is an equal isotropic covariances, γ and ξ are two constants, the default value of γ is 10.

Next, we can therefore rewrite (3) as

$$p_{mn} = \frac{\exp\left[-\left(\frac{1}{2\sigma^2}\mathbf{G}_{mn}^{Global} + \gamma\mathbf{G}_{mn}^{Local} + \xi\mathbf{S}_{mn}^{sift}\right)^2\right]}{\sum_{j=1}^{M}\exp\left[-\left(\frac{1}{2\sigma^2}\mathbf{G}_{jn}^{Global} + \gamma\mathbf{G}_{jn}^{Local} + \xi\mathbf{S}_{jn}^{sift}\right)^2\right] + (2\pi\sigma^2)^{\frac{D}{2}}\frac{\omega}{1-\omega}\frac{M}{N}}. \tag{10}$$

As shown in the function above, (10) differentiate $\mathbf{X}$ and $\mathbf{Y}^w$ by evaluating both global and local geometry structure feature and SIFT feature distance, while (3) concerns the Euclidean distance only. In the case when the geometry structure feature of $\mathbf{X}$ and $\mathbf{Y}^w$ are similar, SIFT feature can still enhance the distinction of the point sets , and provide us a reliable correspondent target point set $\hat{\mathbf{X}}$. After requiring a row-by-row normalization process $p_{mn} \leftarrow p_{mn} / \sum_{j=1}^{M} p_{jn}$ on $\mathbf{P}$, the corresponding target point set $\hat{\mathbf{X}}$ takes the form

$$\hat{\mathbf{X}} = \mathbf{P}\mathbf{X} \tag{11}$$

2.2.2. Transformation Updating

- **Parametric Estimation:** The integrated square error (ISE) or L_2-minimizing estimate(L_2E), a robust estimator is used to minimize the L_2 distance between two models for parametric estimation.
- **Transformation Modeling:** We then define a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$ by choosing a positive definite kernel, here we adopt the Gaussian radial basis function (GRBF), which can be written in the form $\boldsymbol{\mathcal{G}}(\mathbf{y}_i, \mathbf{y}_j) = \exp(-\beta\|\mathbf{y}_i - \mathbf{y}_j\|^2)$, where β is a constant to control the spatial smoothness. According to the representation theorem, the optimal transformation function takes the form $\mathbf{y}^w = \sum_{m=1}^{M} \boldsymbol{\mathcal{G}}(\mathbf{y}, \mathbf{y}_m)\mathbf{c}_m$, where $\mathbf{c}_m$ is a $D \times 1$ coefficient vector. Hence, the minimization over $\mathcal{H}$ boils down to finding a finite M coefficients vector $\mathbf{c}_m$. We formulate the energy function as

$$E(\mathbf{y}^w, \rho^2) = \frac{1}{2^D(\pi\rho)^{\frac{D}{2}}} - \frac{2}{m}\sum_{m=1}^{M}\frac{1}{(2\pi\rho^2)^{\frac{D}{2}}}\exp\left(-\frac{\|\hat{\mathbf{X}}_m - \mathcal{G}_{m,\cdot}\mathbf{C}\|^2}{2\rho^2}\right) + \frac{\lambda}{2}trace(\mathbf{C}^T\boldsymbol{\mathcal{G}}\mathbf{C}), \tag{12}$$

Where ρ^2 controls the convergence of the energy function, it is initialized to 0.05. $\frac{\lambda}{2}trace(\mathbf{C}^T\boldsymbol{\mathcal{G}}\mathbf{C})$ is endowed according to the Tikhonov regularization framework, named global structure preserving term.

Next, we take partial derivative of (12) with respect to coefficient matrix C and obtain the gradient function in the form

$$\frac{\partial E}{\partial \mathbf{C}} = \mathcal{G}^T\left(\frac{2(\mathcal{G}\mathbf{C} - \mathbf{P}\mathbf{X}) \circ (\mathbf{H} \otimes \mathbf{1}_{1\times D})}{n\rho^2(2\pi\rho)^{\frac{D}{2}}} +\right) + \lambda\boldsymbol{\mathcal{G}}\mathbf{C}. \tag{13}$$

where $\mathbf{H} = \exp\{d[(\mathcal{G}\mathbf{C} - \mathbf{P}\mathbf{X})(\mathcal{G}\mathbf{C} - \mathbf{P}\mathbf{X})^T]/2\rho^2\}$ is a $M \times 1$ vector, $d(\cdot)$ denotes the diagonal of a matrix, $\mathbf{1}$ is a $1 \times D$ row vector of all ones. The Hadamard product which can be shown as $(\mathbf{S} \circ \mathbf{T})_{ij} = \mathbf{S}_{ij} \cdot \mathbf{T}_{ij}$ is denoted by $\circ$, and $\otimes$ denotes the Kronecker product. Moreover, non-rigid point set registration has high degrees of flexibility, which meant to say that the optimization process may be trapped by local minima. Therefore, we improve the convergence by adopting deterministic annealing with a manner of coarse-to-fine on the parameter ρ^2. This initializes with a large value for ρ^2 which is progressively decreased by $\rho^2 = \phi\rho^2$, where ρ is a constant. Note that a relatively large ρ makes the annealing scheme slow enough by which the robust result is acquired. And the energy function (12) will converge to a optimal solution eventually.

2.3. Image Transformation Model Estimation

In order to avoid holes and/or overlaps in the output image, a backward approach [43][44] is proposed for building the thin-plate spline (TPS) [45] transformation model. Note that the estimated TPS model from the reference image to the sensed image is the key idea of the backward approach, and can be defined as

$$\mathbf{I}_t(x, y) = \mathbf{I}_s(T(x, y)) \tag{14}$$

where I_t and I_s represent the transformed image and the sensed image, respectively. The size of I_t is same with the reference image I_r, and $T(x, y)$ is the estimated TPS model from the reference image to the sensed image and can be defined as

$$T(\mathbf{Y}^w_x, \mathbf{Y}^w_y) = c_1 + c_x \mathbf{Y}^w_x + c_y \mathbf{Y}^w_y + \sum_{i=1}^{m} \omega_i U(\mathbf{y}^w_i, \mathbf{Y}^w) \tag{15}$$

where $\mathbf{Y}^w_x$ and $\mathbf{Y}^w_y$ indicate the coordinates of feature point set $\mathbf{y}^w$. The kernel function $U(\mathbf{y}^w_i, \mathbf{y}^w_j)$ is defined as $\|\mathbf{y}^w_i - \mathbf{y}^w_j\|^2 \log \|\mathbf{y}^w_i - \mathbf{y}^w_j\|$. Thus, the TPS transformation model $(w_1..., w_m, c_1, c_x, c_y)$ can be estimated by

$$(w_1..., w_m, c_1, c_x, c_y)^T = \begin{bmatrix} \mathbf{K} & \mathbf{P} \\ \mathbf{P}^T & O \end{bmatrix}^{-1} (\mathbf{Y} \mid 0 \quad 0 \quad 0) \tag{16}$$

where ^{T} is the matrix transpose operator and O is a 3 × 3 matrix of zeros. $\mathbf{Y}$ is the $m \times 2$ matrix and indicates the coordinates of feature point set $\mathbf{Y}$. $\mathbf{P}$ is the $m \times 3$ matrix where i^{th} row of $\mathbf{P}$ is (1,$\mathbf{y}^w_{i_x}$,$\mathbf{y}^w_{i_y}$). $\mathbf{K}$ is the $m \times m$ kernel matrix where $\mathbf{K}_{ij} = U(\mathbf{y}^w_i, \mathbf{y}^w_j)$.

2.4. Image resampling and transformation

After the TPS transformation model $(w_1..., w_m, c_1, c_x, c_y)$ is estimated, the transformed image I_t is calculated by the mapping functions constructed during the previous step. The bicubic interpolation algorithm is used in the sensed image I_s on the regular grid. Each pixel from the sensed image can be directly transformed using the estimated mapping functions and the backward approach. The registered image data from the sensed image are determined using the coordinates of the sensed pixel (the same coordinate system as of the reference image) and the inverse of the estimated mapping function. The image interpolation takes place in the sensed image on the regular grid. Neither holes nor overlaps will occur in the output image (i.e., the transformed image I_t) by using the backward approach.

3. Experiments and Results

3.1. Data and Evaluation Criterion

We implemented the proposed method in Matlab, and three series of remote sensing data sets are used to evaluate the performances of the proposed method: (i) UAV image pairs with horizontal viewpoints transformation which contains 100 pairs. (ii) Containing 150 UAV image pairs with vertical viewpoints transformation. (iii) Satellite image registration with horizontal and vertical viewpoints transformation which contains 50 pairs. All of those image pairs were taken at the same time and the images range in overlap is 60-80% . The images have a resolution in the range from 640×450 to 1100×850 pixels. Moreover, compared against five state-of-the-art methods: SIFT[14], SURF [18], RSOC [23], CPD [36], GLMDTPS [41]. All experiments are tested on a PC with 2.20 GHz Intel CPU and 8 GB memory.

A reliable and fair criteria is needed to evaluate the performance of registration approaches because of lacking public different viewpoints remote sensing image registration data sets with desirable ground truth. In this paper, the ground truth is established for matching correctness by checking manually and tried three evaluation methods to evaluate registration approaches. The first one is based on the root of mean square error (RMSE), the second one we measure the median error (MEE), and the third one is based on maximum error (MAE).

More specifically, the ground truth is constructed by the following four steps manually: (1) select at least ten control point pairs in each retinal image pair, (2) compute the transformation via the manual correspondences, (3) transform the matched points in the sensed image to the reference image by the forward spatial transformation, (4) calculate Euclidean distances between the transformed

doi:10.20944/preprints201705.0027.v1

points and the corresponding points in the reference image. Note that we selected the matched points manually using Matlab R2016a and generate the manual ground truth.

3.2. Registration Results on UAV Image with Horizontal Viewpoints Transformation

In the first series of experiments, we test the the performance of our method on the UAV images captured over different province of China by different horizontal viewpoints. The test data sets consist of 100 image pairs and each image has the resolution of 586×452 to 1000×750. We evaluated the performance of our method by all corresponding points which were well-distributed and selected on the easily identified places in I_t and I_r . The registration error was defined as the mean of coordinate differences between the determined corresponding points. Table 1 provides the results of the means and standard deviations of MEE, MAE and RMSE on the whole data sets and compare our method to other five state-of-the-art registration methods, such as SIFT[14], SURF [18], RSOC [23], CPD [36], GLMDTPS [41], our proposed method showed the best performance for all the three criteria. To demonstrate the accurate alignment of our method, all registration examples are shown in Fig. 1.

Table 1. Experimental statistics on UAV images with horizontal viewpoints transformation. Means and standard deviations of MEE, MAE and RMSE for SIFT[14], SURF [18], CPD [36], GLMDTPS [41], RSOC [23] and our method on the whole test data involving 100 UAV image pairs with horizontal viewpoints transformation. Bold indicates the best performance.

	SIFT	**SURF**	**CPD**	**GLMDTPS**	**RSOC**	**OURS**
MEE	14.18264	10.2241	5.8594	5.4353	4.8013	**3.1957**
MAE	16.0080	12.2803	7.5404	7.1459	6.4448	**4.0271**
RMSE	13.5287	7.98374	3.2386	3.0152	2.2737	**1.0171**

3.3. Registration Results on UAV Image with Vertical Viewpoints Transformation

In the second series of experiments, we also test the the performance of our method like section 3.2 on the UAV images captured over different province of China but by different vertical viewpoints. The test data sets consist of 150 image pairs and each image has the resolution of 606×422 to 1150×650. We evaluated the performance of our method by the all corresponding points were well-distributed and selected on the easily identified places in I_t and I_r . The registration error was defined as the mean of coordinate differences between the determined corresponding points. Table 2 provides the results of the means and standard deviations of MEE, MAE and RMSE on the whole data sets and compare our method to other five state-of-the-art registration methods aforementioned where our proposed method also showed the best performance for all the three criteria. Some typical registration examples are shown in Fig. 2.

Table 2. Experimental statistics on UAV images with vertical viewpoints transformation. Means and standard deviations of MEE, MAE and RMSE for SIFT[14], SURF [18], CPD [36], GLMDTPS [41], RSOC [23] and our method on the whole test data involving 150 UAV image pairs with vertical viewpoints transformation. Bold indicates the best performance.

	SIFT	**SURF**	**CPD**	**GLMDTPS**	**RSOC**	**OURS**
MEE	11.9613	8.2523	4.3619	7.6531	5.7844	**3.0021**
MAE	14.2989	9.8411	10.3778	9.1102	7.3585	**3.9188**
RMSE	11.4466	7.0627	7.2645	5.5991	4.1743	**1.4331**

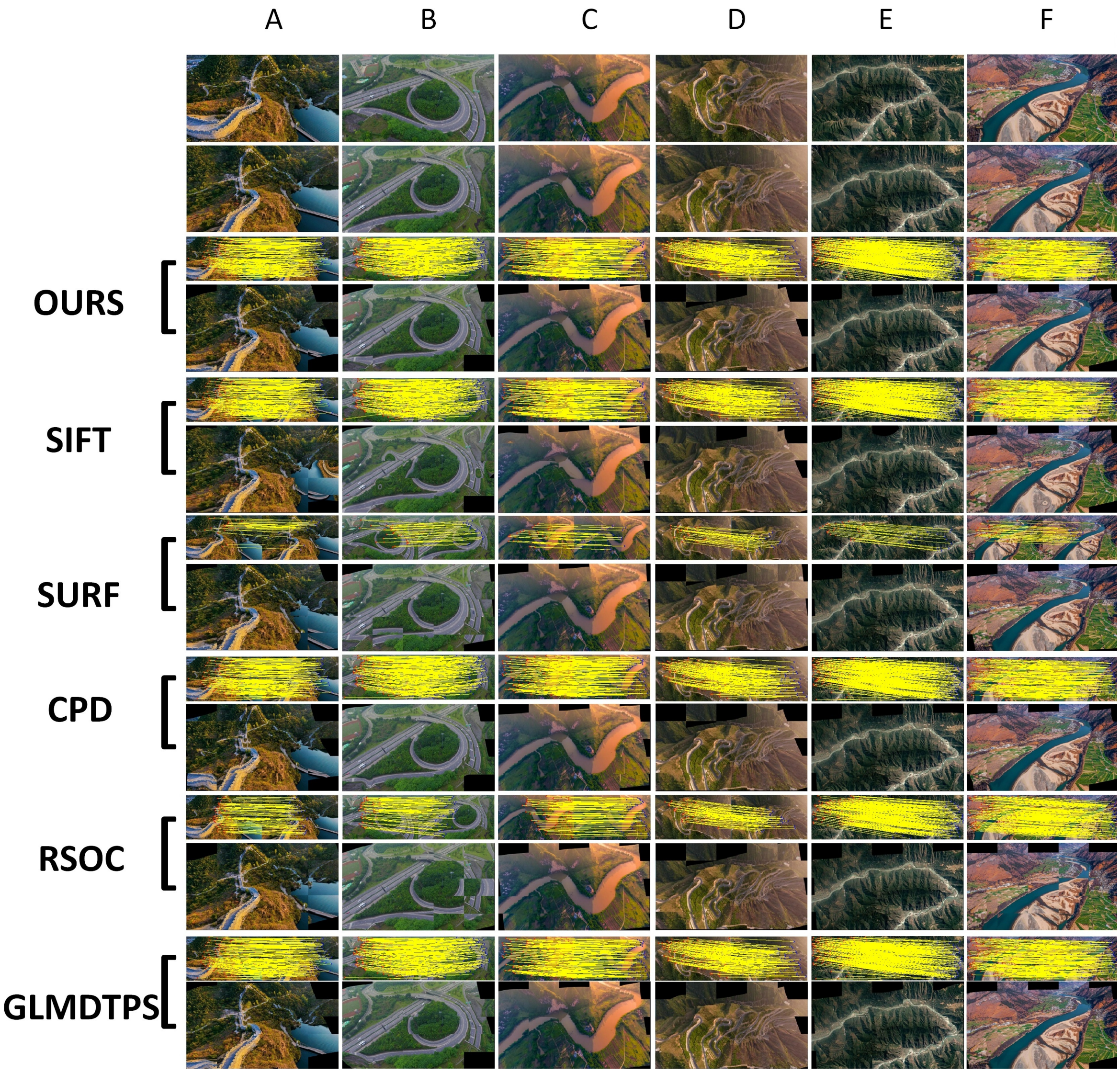

Figure 1. Registration examples on the six typical UAV image pairs. The columns show the registration process for the five UAV image pairs. (A) the Great Wall; (B) ChongQing; (C) GuangXi; (D) XiAn; (E) Jinsha River; (F) the Yangtse River. The first two rows indicate two UAV images (Is and Ir) with different viewpoints. The first row in each method shows the registration results of SIFT feature points, and the second row in each method gives a 5×5 checkerboard for alternately displaying the transformed image It from Is and the image Ir with the different viewpoint as a montage.

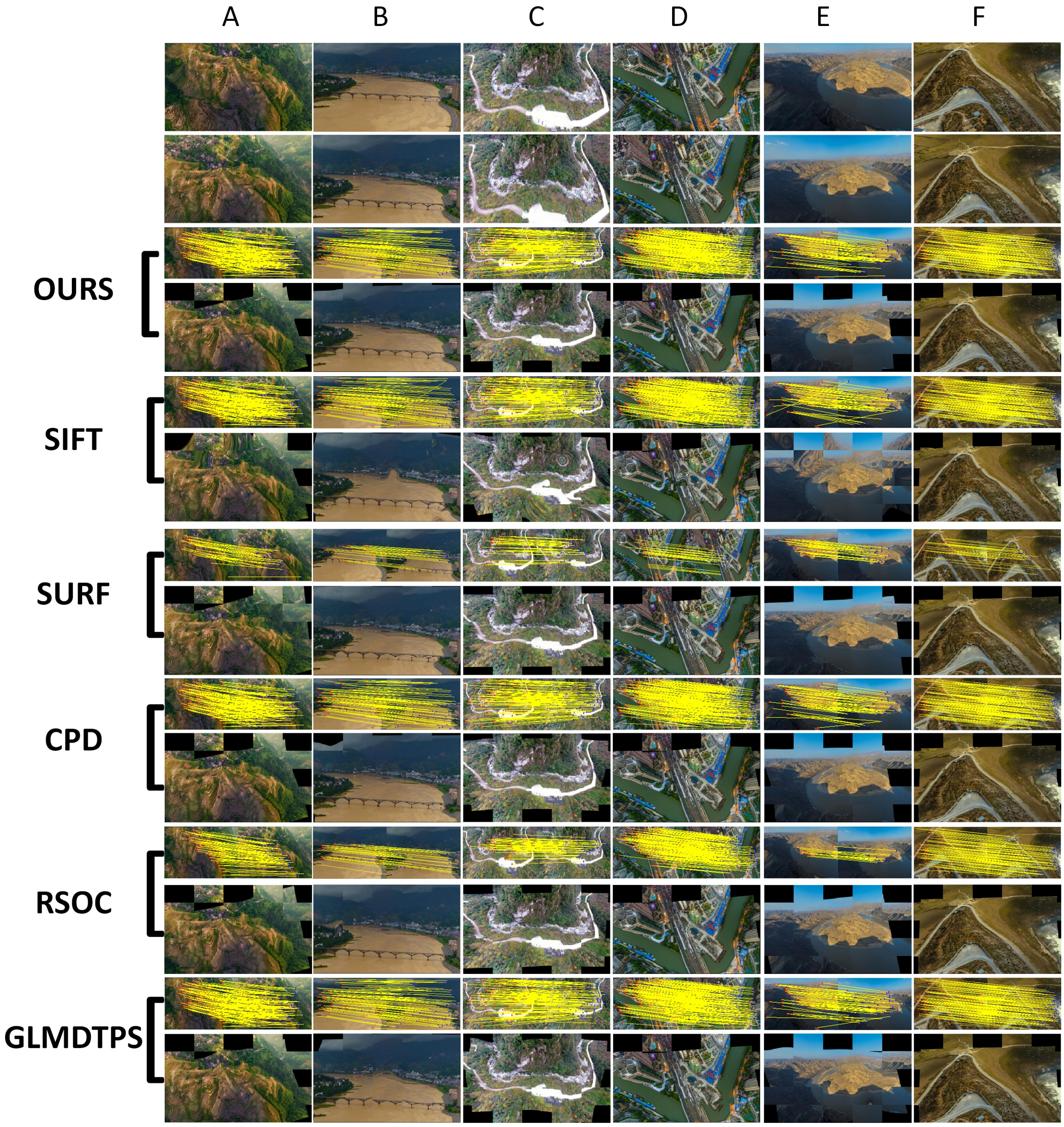

Figure 2. Registration examples on the six typical UAV image pairs. The columns show the registration process for the six UAV image pairs. (A) Longji Rice Terrace; (B) Minqing Bridge; (C) Dongla Grand Canyon; (D) SuZhou River; (E) the Yellow River; (F) Shangri-la. The first two rows indicate two UAV images (I_s and I_r) with different viewpoints. The first row in each method shows the registration results of SIFT feature points, and the second row in each method gives a 5×5 checkerboard for alternately displaying the transformed image I_t from I_s and the image I_r with the different viewpoint as a montage.

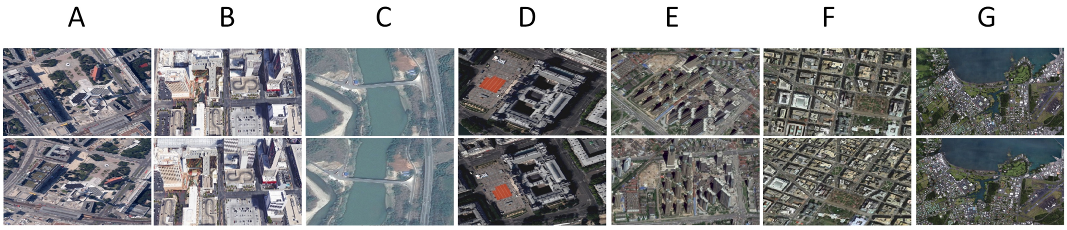

Figure 3. Registration data sets on the seven typical satellite image pairs(i.e. Berlin, Las Vegas, Mekong River, Paris, WuHan, Washington and Hawaii). The first row: the reference image; The second row: the sensed image.

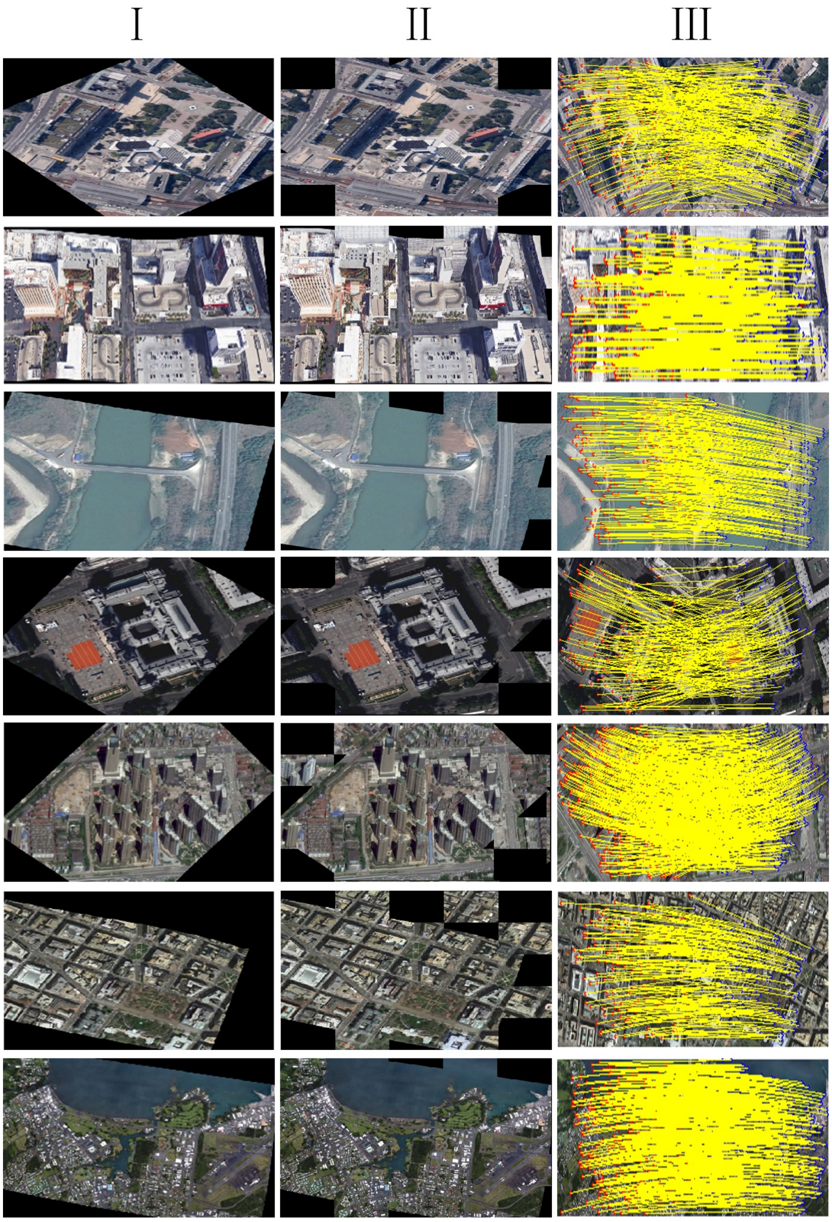

Figure 4. Registration examples on the seven typical satellite image pairs. The rows show the registration process for the seven satellite image pairs. The first column shows the transformed image I_t from I_s, the second row gives a 5×5 checkerboard for alternately displaying I_t and I_r as a montage. The last column shows the registration results of SIFT feature points.

3.4. *Registration Performance on Satellite Image with Horizonal and Vertical Viewpoints Transformation*

In the third series of experiments, we use satellite images obtained from Google earth for evaluation. The test data sets consists of 50 image pairs captured from different places. Each image has the resolution of 586×452, and each image pair (I_s or I_r) involves ground relief variations and imaging viewpoint changes. The ground truth and the registration errors were determined as the same way in section 3.2 and section 3.3. The quantitative registration errors for the proposed method are shown in Table 3, where our method maintains accurate alignments in all experiments. Some typical satellite image data sets (i.e., Berlin, Las Vegas, Mekong River, Paris, WuHan, Washington and Hawaii) and the registration examples are given in Fig. 3 and Fig. 4 respectively.

Table 3. Experimental statistics on satellite image with horizonal and vertical viewpoints transformation, where (A)Berlin; (B)Las Vegas; (C)Mekong River; (D)Paris; (E)WuHan; (F)Washington; (G)Hawaii. Means and standard deviations of MEE, MAE and RMSE for seven typical image pairs with horizonal and vertical viewpoints transformation are shown.

	A	**B**	**C**	**D**	**E**	**F**	**G**	**Average**
MEE	1.8562	2.5498	2.5171	1.8061	2.1429	2.0098	2.0717	2.1362
MAE	2.1458	4.0208	2.9097	1.9317	2.5778	2.3542	2.2958	2.6051
RMSE	0.8358	1.3963	1.9556	1.4272	0.7743	1.4342	0.3171	1.1628

4. Conclusion

In this paper, we proposed an accurate method based on local invariant features and geometric structure descriptors. This approach contains SIFT feature extraction, feature point sets registration based on improved SIFT algorithm via geometric structure constraint and image transformation by estimating non-rigid image transformation model. The main contributions of this work are considered as the proposed four-step approach which solved the remote sensing image registration problems in the moderate ground relief variations and imaging viewpoint changes. Experiments on both UAV images and satellite images from different viewpoints demonstrate that our method shows best registration performances against five state-of-the-art methods .

Acknowledgments

The authors wish to thank David G. Lowe, Herbert Bay, Andriy Myronenko, Zhaoxia Liu and Yang Yang for providing their implementation source codes and test data sets. This greatly facilitated the comparison experiments. This work was supported by (i) National Nature Science Foundation of China [41661080]; (ii) Key Scientific Research Project of Education Department of Yunnan Province [2015Z069]; (iii) Scientific Research Foundation of Yunnan Provincial Department of Education; (iv) Doctoral Scientific Research Foundation of Yunnan Normal University [01000205020503065].

Conflict of interest

The authors declare that they have no conflicts of interest in the research.

References

1. Liu, Z.; An, J.; Jing, Y. A simple and robust feature point matching algorithm based on restricted spatial or derconstraints for aerial image registration. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 514-527.
2. Zitov, A. B.; Flusser, J. Shape matching and object recognition using shape contexts. *Image Vision Compute.* **2003**, *21*, 977-1100.
3. Brown, L. G. A survey of image registration techniques. *ACM Compute Surv.* **1992**, *24*, 325-376.
4. Maintz, J. B. A.; Viergever, M. A. A survey of medical image registration. *Med. Image Anal.* **1998**, *2*, 1-36.

5. Wang, X.; Li, Y.; Wei, H. and Liu, F. An asift-based local registration method for satellite imagery. *Remote Sens.* **2015**, *7*, 7044-7061.
6. Liu S.; Tong X.; Chen J.; et al. A Linear Feature-Based Approach for the Registration of Unmanned Aerial Vehicle Remotely-Sensed Images and Airborne LiDAR Data. *Remote Sens.* **2016**, *8*,82.
7. Chen, J.; Tian, J.; Lee, N. and Zheng, J. A partial intensity invariant feature descriptor for multimodal retinal image registration. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1707-1718.
8. Goncalves, H.; Goncalves, J. A.; Cortereal, L. Automatic image registration based on correlation and Hough transform. *SPIE Remote Sensing. International Society for Optics and Photonics.* **2008**, *7109*, 71090J-71090J-12.
9. Tong, X.; Ye, Z.; Xu, Y. and Liu, S. A Novel Subpixel Phase Correlation Method Using Singular Value Decomposition and Unified Random Sample Consensus. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4143-4156.
10. Xu, M.; Varshney, P. K.; Niu, R. A Subspace Method for Fourier Based Image Registration. *Signals, Systems and Computers, 2006. ACSSC '06. Fortieth Asilomar Conference.* **2006**, *6*, 425-429.
11. Chen, H. M.; Varshney, P. K.; Arora, M. K. Performance of mutual information similarity measure for registration of multitemporal remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2445-2454.
12. Song, Z .; Zhou, S.; Guan, J. A novel image registration algorithm for remote sensing under affine transformation. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4895-4912.
13. Jensen, J. R. Introductory digital image processing. *Prentice Hall: Upper Saddle River, NJ, USA.* **2004**, *2*, 382-382.
14. Lowe, D. G. Object recognition from local scale-invariant features. *The Proceedings of the Seventh IEEE International Conference on Computer Vision. IEEE.* **1999**, *2*, 1150.
15. Lowe, D. G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Compute Vision* **2004**, *60*, 91-110.
16. Lowe, D. G. Method and apparatus for identifying scale invariant features in an image and use of same for locating an object in an image. US, US6711293[P]. 2004.
17. Goncalves, H.; Corte-Real, L.; Goncalves, J. A. Automatic Image Registration Through Image Segmentation and SIFT. *EEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2589-2600.
18. Bay, H.; Ess, A.; Tuytelaars, T. and Van Gool, L. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346-359.
19. Harris,C. A combined corner and edge detector. Proc Alvey Vision Conf. **1988** ,*1988* ,147-151.
20. Cheng, L.; Gong, J.; Yang, X.; Fan, C. Robust Affine Invariant Feature Extraction for Image Matching. *IEEE Geosci. Remote Sens. Lett* **2008** , *5*,246-250.
21. Brook, A.; Bendor, E. Automatic registration of airborne and spaceborne images by topology map matching with surf processor algorithm.*Remote Sens.* **2011** , *3*, 65-82.
22. Sima, A. A.; Buckley, S. J. Optimizing sift for matching of short wave infrared and visible wavelength images.*Remote Sens.* **2013** , *5*, 2037-2056.
23. Liu, Z.; An, J.; Jing, Y. A Simple and Robust Feature Point Matching Algorithm Based on Restricted Spatial Order Constraints for Aerial Image Registration. *IEEE Trans. Geosci. Remote Sens.* **2012** , *50*, 514-527.
24. Fan, B.; Wu, F.; Hu, Z. Rotationally Invariant Descriptors Using Intensity Order Pooling. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011** , *34*, 2031-45.
25. Mikolajczyk, K.; Schmid, C. A performance evaluation of local descriptors. *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. IEEE.* **2003** , II-257-II-263 vol.2.
26. Li, Q.; Wang, G.; Liu, J.; Chen, S. Robust Scale-Invariant Feature Matching for Remote Sensing Image Registration. *IEEE Geosci. Remote Sens. Lett.* **2009** , *6*, 287-291.
27. Sedaghat, A.; Mokhtarzade, M.; Ebadi, H. Uniform Robust Scale-Invariant Feature Matching for Optical Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2011** , *49*, 4516-4527.
28. Ke, Y.; Sukthankar, R. 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. IEEE.* **2004** , *2* , 506-513.
29. Dellinger, F.; Delon, J.; Gousseau, Y.; Michel, J. SAR-SIFT: A SIFT-Like Algorithm for SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2015** , *53*, 453-466.
30. Sedaghat, A.; Ebadi, H. Remote Sensing Image Matching Based on Adaptive Binning SIFT Descriptor. *IEEE Trans. Geosci. Remote Sens.* **2015** , *53*, 5283-5293.

31. Liu, F.; Bi, F.; Chen, L.; Shi, H. Feature-Area Optimization: A Novel SAR Image Registration Method. *IEEE Geosci. Remote Sens. Lett.* **2016** , *13*, 242-246.
32. Goncalves, H.; Corte-Real, L.; Goncalves, J A. Automatic Image Registration Through Image Segmentation and SIFT. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2589-2600.
33. Gong, M.; Zhao, S.; Jiao, L.; Tian, D. A Novel Coarse-to-Fine Scheme for Automatic Image Registration Based on SIFT and Mutual Information. *IEEE Trans. Geosci. Remote Sens.* **2014** , *52*, 4328-4338.
34. Ma, J.; Zhao, J.; Tian, J.; Yuille, A. L. & Tu, Z. Robust point matching via vector field consensus. *IEEE Trans. Image Process.* **2014**, *23*, 1706-1721.
35. Fischler M A, Bolles R C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the Acm.* **1980** , *24*, 381-39
36. Myronenko, A.; Song, X. Point set registration: coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009** , *32(12)*, 2262-2275.
37. Liu, H.; Yan, S. Common visual pattern discovery via spatially coherent correspondences. *IEEE conf. on Comput. Vis. and Pattern Recognit.* **2010**, 1609-1616.
38. Zhang, Z.; Yang, M.Y. Zhou, M.; Zeng, X.Z. Simultaneous remote sensing image classification and annotation based on the spatial coherent topic model. *IGARSS 2014 - 2014 IEEE International Geoscience and Remote Sensing Symposium. IEEE.* **2014** , 1698-1701.
39. Ma, J.; Qiu, W.; Zhao, J.; Ma, Y. Robust, Estimation of Transformation for Non-Rigid Registration. *IEEE Trans. Signal Process.* **2015** , *63*, 1115-1129.
40. Ma, J.; Zhou, H.; Zhao, J.; Gao, Y.; Jiang, J.J.; Tian,J.W. Robust Feature Matching for Remote Sensing Image Registration via Locally Linear Transforming. *IEEE Trans. Geosci. Remote Sens.* **2015** , *53*, 6469-6481.
41. Yang, Y.; Ong S H, Foong, K. W. C. A robust global and local mixture distance based non-rigid point set registration. Pattern Recognit. **2015** , *48*, 156-173.
42. Belongie, S.; Malik, J.; Puzicha, J. Shape matching and object recognition using shape contexts. *IEEE T Pattern Ana.* **2002**, *24*, 509-522.
43. Zitova, B.; Flusser, J. Image registration methods: a survey. *Image vision comput.* **2003** , *21*, 977-1000.
44. Ji, S.; Peng, S. Terminal perturbation method for the backward approach to continuous time mean-variance portfolio selection. *Stochastic Process. Appl.* **2008** , *118*, 952-967.
45. Bookstein, F. L. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989** , *11* , 567-585.