



Robust feature matching via Gaussian field criterion for remote sensing image registration

Qing Ma¹ · Xu Du¹ · Jiahao Wang² · Yong Ma² · Jiayi Ma² 

Received: 10 December 2017 / Accepted: 21 February 2018 / Published online: 2 April 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Feature matching, which refers to establishing reliable feature correspondences between two images of the same scene, is a critical prerequisite in a wide range of remote sensing tasks including environment monitoring, multispectral image fusion, image mosaic, change detection, map updating. In this paper, we propose a method for robust feature matching and apply it to the problem of remote sensing image registration. We start by creating a set of putative feature matches which can contain a number of unknown false matches, and then focus on mismatch removal. This is formulated as a robust regression problem, and we customize a robust estimator, namely the Gaussian field criterion, to solve it. The robust criterion can handle both linear and nonlinear image transformations. In the linear case, we use a general homography to model the transformation, while in the nonlinear case, the non-rigid functions located in a reproducing kernel Hilbert space are considered, and a regularization term is added to the objective function to ensure its well-posedness. Moreover, we apply a sparse approximation to the non-rigid transformation and reduce the computational complexity from cubic to linear. Extensive experiments on various natural and remote sensing images show the effectiveness of our approach, which is able to yield superior results compared to other state-of-the-art methods.

Keywords Feature matching · Image registration · Remote sensing · Gaussian field · Robust estimation

1 Introduction

Image registration is a vital prerequisite in many applications for remote image analysis [1–9]. The primary goal of image registration is to spatially align two or more different images of the same scene (i.e., the reference and sensed

images) taken at different times, from different viewpoints, or by different sensors [10].

Over the years, a number of methods have been proposed to deal with the problem of automated remote image registration [11, 12]. Generally, these methods can be divided into two categories: methods based on pixel intensities and methods based on image features [10, 12]. Methods based on pixel intensities use original pixel intensities to find the matching information in the overlap area of two images, while methods based on image features seek the correspondence between two sets of features based on descriptors and/or using certain spatial constraints. Methods based on pixel intensities are suitable when the transformations between images are not so complex. These methods are sensitive to noise, illumination variance, and difference of sensor types. Besides, they also suffer from heavy computational complexities. On the contrary, methods based on image features are relatively more robust, as features are not so easy to be influenced by the changes mentioned above, which allow them to handle more complex image distortions. The features used for

✉ Jiayi Ma
jiayima@whu.edu.cn

Qing Ma
qingma2017@mails.ccnu.edu.cn

Xu Du
duxu@mail.ccnu.edu.cn

Jiahao Wang
wangjiahao@whu.edu.cn

Yong Ma
mayong@whu.edu.cn

¹ National Engineering Research Center for E-Learning, Central China Normal University, Wuhan 430079, China

² Electronic Information School, Wuhan University, Wuhan 430072, China

image registration are typical point features obtained by feature detectors such as scale-invariant feature transform (SIFT) [13], as they are more general and easier to be extracted. In this paper, we focus on point-based feature registration techniques.

The point-based matching problem is typically solved by using a two-step strategy [14]. In the first stage, a set of putative matches is constructed based on a similarity constraint requiring that two feature points can be matched if they have similar descriptors. In this process, the scale of possible matches is largely reduced; however, the putative matches established by only local feature descriptors are unreliable due to viewpoint changes, repeated patterns, as well as occlusions. Therefore, in the second step, the geometric constraint is further adopted to remove the false matches and the matching task boils down to determining the correctness of each putative match.

Although quite a lot of point-based registration methods have been investigated in the past decades, it is still a challenging task to design an efficient strategy when dealing with real-world remote sensing image registration problems. First, the image transformation model is often unknown in advance and varies with respect to different types of remote sensing scenes, making it difficult to design a general algorithm. Second, a large number of false matches will exist due to low overlaps, repeated patterns, or local distortions of remote sensing images, requiring high robustness of the matching model. Third, high computational complexity is often incurred, especially in case of non-rigid transformation, limiting its applicability in handling real-time problems or matching high-resolution remote sensing images.

In view of the above problems, we propose an efficient method for robust feature matching of remote sensing images based on Gaussian field criterion. The mismatch removal is formulated as a robust regression problem, and we customize the Gaussian field criterion to solve it to achieve a robust estimation. We model the image transformation with both linear and nonlinear functions to handle various matching problems, and a regularization term is imposed in the nonlinear case to ensure well-posedness. In addition, we apply a sparse approximation to the transformation to significantly reduce the computational complexity.

Our contribution in this paper includes three aspects. Firstly, we introduce the Gaussian field criterion and analyze its robustness both in theory and in experiment, which provides support for why it can resist outliers. Secondly, we customize the Gaussian field criterion to address the feature matching problem for remote sensing image registration, which can efficiently remove false matches from a set of putative matches with even up to 80% outliers. Thirdly, we adopt both linear and nonlinear transformation

model together with a sparse approximation, which can deal with more general matching problem in low computational complexity, especially in the nonlinear situation.

The rest of the paper is organized as follows: Section 2 overviews the previous work related to image registration, especially in the context of remote sensing image registration. In Sect. 3, we propose the Gaussian field criterion and customize it to solving the linear and nonlinear feature matching problem. Subsequently, we present the fast implementation of the proposed method and discuss the computational complexity. In Sect. 4, we evaluate the performance of our method on various natural as well as remote sensing image pairs with comparisons to several state-of-the-art approaches, followed by some concluding remarks in Sect. 5.

2 Related work

Image registration benefits many research fields, including computer vision [15–17], pattern recognition [18–21], medical imaging [22, 23], and remote sensing [24–28]. Here we briefly overview two major categories of registration methods (i.e., methods based on intensities and methods based on features) that are most relevant to our approach, particularly in the context of remote sensing registration.

2.1 Methods based on pixel intensities

Methods based on pixel intensities typically use windows of certain size for the correspondence estimation without detecting salient objects in images, such as features. The most representative methods that fall in this category are correlation-based methods [29], Fourier methods [30], and mutual information methods [31].

Classical correlation-based methods, such as cross-correlation [29], compute the similarities of windows in two images. They deal directly with pixel intensities, without any structural analysis. There have been some of the correlation-based methods using maxima of wavelet coefficients in remote sensing image registration as in [32]. Correlation-based methods are often used in real-time applications because of their easy implementation in hardware. However, these methods have some obvious disadvantages when the images in windows suffer from complex transformations or contain relatively smooth areas without prominent details.

Fourier methods exploit the Fourier representation of images in the frequency domain. One representative is the phase correlation method based on the Fourier shift theorem, which was originally proposed to deal with translated images and later extended to account for rotation and

scaling [30]. The applications to remote sensing are described in [33], which has shown that the images can be aligned using a Fourier–Mellin transform together with the phase correlation. Compared to the correlation-based methods, the Fourier methods have the advantages of computational efficiency and have shown strong robustness against varying imaging conditions and frequency-dependent noise.

Mutual information, which measures the statistical dependency between two sets of data, comes from the information theory. Mutual information methods take such information as the metric to maximize the dependency between the given images. These methods are particularly suitable for multimodal registration; for example, in remote sensing applications the images are often captured by different modalities of sensors to exploit complementary information [34]. Rather than operating directly on the image intensities, a mutual information method that operates on the extracted features such as points of the area borders has also been developed [31]. In addition, the mutual information methods are also frequently adopted in medical imaging to compare the anatomical and functional images of the patient's body, leading to a diagnosis accordingly [31].

2.2 Methods based on image features

The second category of image registration methods relies on features, namely visually salient structures in images. It is usually assumed in these methods that the features extracted can be represented as a set of spatial points, called control points in the literature [12].

A popular strategy for solving the matching problem involves two steps [14]: (i) computing a set of putative correspondences and (ii) then removing the outliers via geometrical constraints. Putative correspondence instances are obtained in the first step by pruning the set of all possible point correspondences. This scenario is achieved by computing feature descriptors at the points and eliminating the matches between points whose descriptors are excessively dissimilar. Lowe [13] proposed the SIFT descriptor with a distance ratio method that compares the ratio between the nearest and next-nearest neighbors against a predefined threshold to filter out unstable matches. Guo and Cao [35] proposed a triangle constraint, which can produce better putative correspondences in terms of quantity and accuracy compared with the distance ratio in [13]. Pele and Michael [36] applied the earth mover's distance to replace the Euclidean distance in [13] to measure the similarity between descriptors and improve the matching accuracy. Li et al. [37] further refined the descriptor of SIFT to overcome the difference in the gradient intensity and orientation between remote

image pairs. In addition, Hu et al. [38] proposed the local selection of a suitable descriptor for each feature point instead of employing a global descriptor during putative correspondence construction. A cascade scheme has been suggested to prevent the loss of true matches, which can significantly enhance the correspondence number [39, 40]. Although there have been various sophisticated approaches for putative match construction, the use of only local appearance features will inevitably result in a lot of false matches. In the second step, robust estimators based on geometrical constraints are used to detect and remove the outliers.

To efficiently remove mismatches, the traditional resampling methods such as random sample consensus (RANSAC) [41] and its variants [42, 43] have achieved great success in many situations. However, they rely on a predefined parametric model, which become less efficient when the underlying image transformation is non-rigid. In addition, they also tend to severely degrade if the outlier proportion becomes large. To address these issues, several nonparametric interpolation methods have recently been introduced, such as identifying correspondence function (ICF) [44], graph shift (GS) [45], vector field consensus (VFC) [14] and locally linear transforming (LLT) [10]. These methods usually interpolate the underlying image transformation with kernel functions using unsupervised learning and are able to handle complex non-rigid deformation. Specifically, by using support vector regression, the ICF method learns a correspondence function pair mapping feature points across two images and then rejects the outliers by checking whether they are consistent with the estimated correspondence functions. The GS method constructs an affinity graph for the correspondences based on the spectral matching method, and the maximal clique of the graph is viewed as spatially coherent correspondences. The VFC method learns a smooth field to fit the potential inliers, while the LLT method recovers a smoothness transformation according to a local linear constraint, and both can resist a great quantity of outliers. Specifically, in the remote sensing community, Li et al. [25] proposed a robust method based on support-line voting and affine-invariant ratios, which has shown large improvements compared with RANSAC. Yao et al. [46] introduced a Harris corner matching method based on the quasi-homography transform (QHT) and self-adaptive window, where the QHT can reduce the search range effectively for match candidates. In addition, Yu et al. [47] presented a method based on LLT and a manifold regularization (MR) [48] for feature matching of unmanned aerial vehicle (UAV) images.

The feature matching can also be formulated in terms of a correspondence matrix between feature points together with a parametric, or nonparametric, geometric constraint.

Specifically, Boughorbel et al. [49] introduced the Gaussian fields into 3D rigid shape registration, which was later generalized to the non-rigid setting in [50]. Based on this concept [51], further promoted the matching performance by constructing a context-aware representation for assignment initialization. During the manuscript preparing, the authors in [51] introduced a mismatch removal method named Gaussian field consensus [52] for nonparametric feature matching based on a locally linear constraint similar to LLT [10]. In this paper, we formulate the parametric and nonparametric models in a general framework based on the Gaussian field criterion for remote sensing image matching. It is robust and general and is able to handle both linear and nonlinear distortions in low computational complexity.

3 Method

This section describes our method for robust feature matching. To this end, we first create a set of putative matches by considering all possible matches and pruning matches whose feature descriptors are sufficiently different. Fortunately, this can be achieved by using existing well-designed local feature descriptors such as SIFT [13]. Next, we will introduce a robust estimator and customize it to further remove false matches from the putative set based on a global geometric constraint.

3.1 Gaussian field criterion

Many problems in applied mathematics, such as the heat equation, vortex methods, tomography and nonparametric statistics, involve the Gauss transform [53]

$$G_\sigma \mathbf{f}(\mathbf{x}) = \int_{\mathcal{H}} e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{\sigma^2}} \mathbf{f}(\mathbf{y}) d\mathbf{y}, \quad (\sigma > 0) \tag{1}$$

of a function \mathbf{f} defined on $\mathcal{H} \subset \mathbb{R}^D$ with σ being a range parameter. In practical problem, the $G_\sigma \mathbf{f}$ is usually discretized for numerical computation. Specifically, given the values of \mathbf{f} at a set of points $\{\mathbf{s}_j\}_{j=1}^N \subset \mathbb{R}^D$, the integral (1) could be approximated by the following discrete Gauss transform

$$G(\mathbf{x}) = \sum_{j=1}^N q_j e^{-\frac{\|\mathbf{x}-\mathbf{s}_j\|^2}{\sigma^2}}, \tag{2}$$

where q_j depending on $\mathbf{f}(\mathbf{s}_j)$ is a coefficient. Equation (2) is also called *Gaussian field* due to sources of strengths $\{q_j\}_{j=1}^N$ at the point set $\{\mathbf{s}_j\}_{j=1}^N$, evaluated at the “target” \mathbf{x} .

The Gaussian field is a sum of Gaussian distances which is robust to noise and outliers. (We will discuss this property later.) It has been applied to the shape registration

problem and demonstrated promising results [49, 50]. In the following, we will customize it to solve the feature matching problem and show its robustness both in theory and in experiment.

3.2 Problem formulation

Suppose we have obtained a set of putative feature matches $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$ with \mathbf{x}_i and \mathbf{y}_i being spatial positions of feature points extracted from two images. The putative set may be contaminated by some unknown false matches, and our goal is to distinguish those outliers to establish accurate matches. Typically, the underlying inliers obey a global geometric constraint; for example, there exists a spatial transformation \mathbf{f} satisfying that $\forall i \in \mathbb{N}_N, \mathbf{y}_i = \mathbf{f}(\mathbf{x}_i)$ if $(\mathbf{x}_i, \mathbf{y}_i)$ is an inlier. Thus, the matching problem boils down to seeking the solution of \mathbf{f} and a robust estimation is desirable due to the existence of outliers.

Intuitively, considering the noiseless case, the optimal solution \mathbf{f} is the one leading to the maximum point-to-point overlap between $\{\mathbf{f}(\mathbf{x}_i)\}_{i=1}^N$ and $\{\mathbf{y}_i\}_{i=1}^N$. This can be formulated by using a Boolean operator that “counts” overlapping point number under the transformation \mathbf{f} :

$$\mathcal{E}(\mathbf{f}) = \sum_{i=1}^N \delta(d(\mathbf{f}(\mathbf{x}_i), \mathbf{y}_i)), \quad \delta(t) = \begin{cases} 1, & t = 0 \\ 0, & t \neq 0 \end{cases}, \tag{3}$$

where $d(\mathbf{x}, \mathbf{y})$ is a distance metric such as Euclidean distance. Equation (3) is a combinational optimization problem which is difficult to obtain the optimal solution. Rather than using the discontinuous indicator function δ , we consider the Gaussian field in Eq. (2) and obtain the Gaussian field criterion:

$$G(\mathbf{f}) = - \sum_{i=1}^N e^{-\frac{d^2(\mathbf{f}(\mathbf{x}_i), \mathbf{y}_i)}{\sigma^2}}. \tag{4}$$

Now we consider the robustness of the Gaussian field criterion (3). On the one hand, the point position noise can be dealt with by tuning the range parameter σ to the noise variance. On the other hand, the Gaussian distance is inherently robust to outliers, which can be seen from penalty curves of L_2 and Gaussian distance in Fig. 1. Specifically,

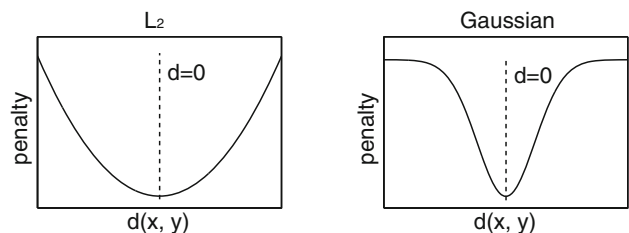


Fig. 1 The penalty curves of L_2 and Gaussian distances

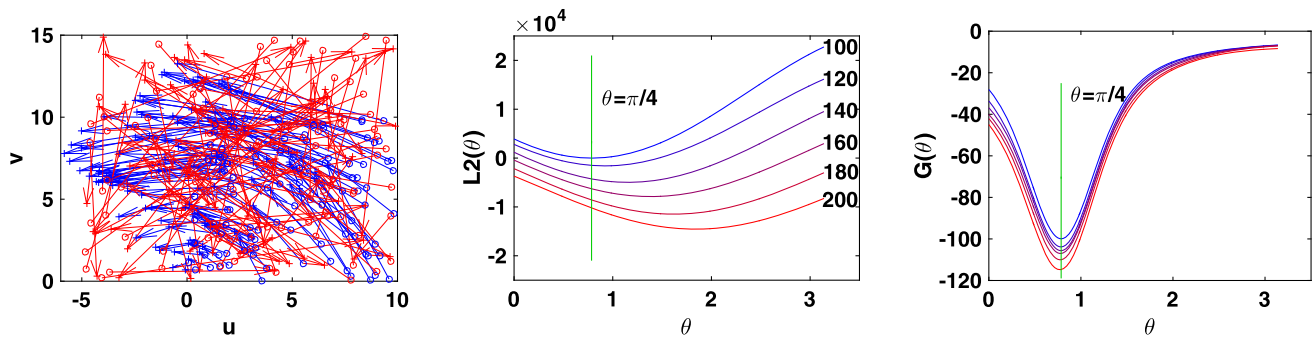


Fig. 2 Robustness comparison between the L_2 and Gaussian distances. Left: data samples containing 100 inliers (denoted in blue) with Gaussian noise and 100 random outliers (denoted in red), where the “o” and “+”, respectively, denote the point sets $\{\mathbf{x}_i\}_{i=1}^N$ and $\{\mathbf{y}_i\}_{i=1}^N$, and the spatial transformation is a rotation (i.e., $\mathbf{y}_i = R(\theta)\mathbf{x}_i$)

the L_2 loss, i.e., $L_2(\mathbf{f}) = \sum_{i=1}^N d^2(\mathbf{f}(\mathbf{x}_i), \mathbf{y}_i)$, has quadratic penalty curve and resists assigning large distance $d^2(\mathbf{f}(\mathbf{x}_i), \mathbf{y}_i)$ to any matches and hence tends to be biased by outliers. In contrast, the Gaussian distance could be approximately seen as a truncated L_2 loss, which can bear large distance without paying a too high penalty. In addition, the Gaussian distance is computationally efficient, since it can be simplified into its corresponding logarithmic form.

We further give some intuition on the robustness through the experiment in Fig. 2. The two point sets are related by a rotation transformation, and we are required to seek the optimal rotation angle θ . From the results, we see that the L_2 loss works well if the data samples do not involve outliers; however, its global optimum becomes steadily worse as the amount of outliers increases. In contrast, the Gaussian field criterion can always produce a global optimum near the ground truth $\theta = \pi/4$ even half of the data are outliers.¹

To solve the Gaussian field criterion in Eq. (4), we need to model the transformation \mathbf{f} . As can be seen from Eq. (4), the Gaussian field criterion is quite general, which does not rely on any specific transformation model. Next, we model the transformation with homography and non-rigid model for linear and nonlinear matching, respectively.

3.3 Matching with linear transformation model

For image pairs related by a linear transformation model, we consider the general projection transformation which is

with $\theta = \pi/4$ as the ground truth. Middle and right: the curves of L_2 and Gaussian distances with respect to the rotation angle θ . Initially, the test data contain only 100 inliers, and then we vary it by always adding 20 new outliers until 100 outliers

characterized by a 3×3 homography matrix \mathbf{H} . Specifically, we use the homogeneous coordinate for each feature point, e.g., $\mathbf{x}_i = (\mathbf{x}_i^u, \mathbf{x}_i^v, 1)^T$ and $\mathbf{y}_i = (\mathbf{y}_i^u, \mathbf{y}_i^v, 1)^T$. Thus, the transformation has the following form:

$$\mathbf{y}_i = \mathbf{H}\mathbf{x}_i. \tag{5}$$

Due to the use of homogeneous coordinate, the 3D vectors \mathbf{y}_i and $\mathbf{H}\mathbf{x}_i$ are not equal, which have the same direction but may differ in magnitude by a nonzero scale factor. Thus, the equation can be expressed in terms of vector cross-product, i.e., $\mathbf{y}_i \times \mathbf{H}\mathbf{x}_i = \mathbf{0}$. By denoting \mathbf{h}^j the j th row of \mathbf{H} , we obtain the following equation:

$$\mathbf{y}_i \times \mathbf{H}\mathbf{x}_i = \mathbf{y}_i \times \begin{pmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \mathbf{h}^3 \end{pmatrix} \mathbf{x}_i = \begin{pmatrix} \mathbf{y}_i^v \mathbf{h}^3 \mathbf{x}_i - \mathbf{h}^2 \mathbf{x}_i \\ \mathbf{h}^1 \mathbf{x}_i - \mathbf{y}_i^u \mathbf{h}^3 \mathbf{x}_i \\ \mathbf{y}_i^u \mathbf{h}^2 \mathbf{x}_i - \mathbf{y}_i^v \mathbf{h}^1 \mathbf{x}_i \end{pmatrix}. \tag{6}$$

Therefore, the equation $\mathbf{y}_i \times \mathbf{H}\mathbf{x}_i = \mathbf{0}$ has the follow form:

$$\begin{pmatrix} \mathbf{y}_i^v \mathbf{h}^3 \mathbf{x}_i - \mathbf{h}^2 \mathbf{x}_i \\ \mathbf{h}^1 \mathbf{x}_i - \mathbf{y}_i^u \mathbf{h}^3 \mathbf{x}_i \\ \mathbf{y}_i^u \mathbf{h}^2 \mathbf{x}_i - \mathbf{y}_i^v \mathbf{h}^1 \mathbf{x}_i \end{pmatrix} = \begin{bmatrix} \mathbf{0}^T & -\mathbf{x}_i^T & \mathbf{y}_i^v \mathbf{x}_i^T \\ \mathbf{x}_i^T & \mathbf{0}^T & -\mathbf{y}_i^u \mathbf{x}_i^T \\ -\mathbf{y}_i^v \mathbf{x}_i^T & \mathbf{y}_i^u \mathbf{x}_i^T & \mathbf{0}^T \end{bmatrix} \mathbf{h} = \mathbf{0}, \tag{7}$$

where $\mathbf{h} = (\mathbf{h}^1, \mathbf{h}^2, \mathbf{h}^3)^T$ is a nine-element vector made up of the entries of \mathbf{H} in row-major order.

Note that Eq. (7) contains three equations, only two of which are linearly independent. Thus, a point correspondence produces only two equations; here, we utilize the former two and obtain the following equation:

$$\mathbf{W}_i \mathbf{h} = \mathbf{0}, \tag{8}$$

where the coefficient matrix \mathbf{W}_i is defined as:

¹ Note that the Gaussian criteria are not fully robust to noise and outliers. It is essentially a low-pass filter that filters out high-frequency components (usually noise) based on the penalty factor. But it is ineffective in cases of salt and pepper noise characterized outliers.

$$W_i = \begin{bmatrix} \mathbf{x}_i^u & \mathbf{x}_i^v & 1 & 0 & 0 & 0 & -\mathbf{y}_i^u \mathbf{x}_i^u & -\mathbf{y}_i^u \mathbf{x}_i^v & -\mathbf{y}_i^u \\ 0 & 0 & 0 & \mathbf{x}_i^u & \mathbf{x}_i^v & 1 & -\mathbf{y}_i^v \mathbf{x}_i^u & -\mathbf{y}_i^v \mathbf{x}_i^v & -\mathbf{y}_i^v \end{bmatrix}. \tag{9}$$

Therefore, the Gaussian field criterion in Eq. (4) for linear matching becomes:

$$G(\mathbf{h}) = - \sum_{i=1}^N e^{-\frac{\|\mathbf{W}_i \mathbf{h}\|^2}{\sigma^2}}. \tag{10}$$

Clearly, the optimal solution of the above objective function is $\mathbf{h} = \mathbf{0}$, and it is of no interest to us. Actually, the homography \mathbf{H} is a homogeneous matrix, consisting of nine elements, but it has only eight degrees of freedom. Without loss of generality, we fix the last element of \mathbf{h} to 1. By defining $\tilde{\mathbf{h}}$ as an eight-element vector (i.e., the first eight elements of the original \mathbf{h}) and $\tilde{\mathbf{W}}_i$ as follows:

$$\tilde{\mathbf{W}}_i = \begin{bmatrix} \mathbf{x}_i^u & \mathbf{x}_i^v & 1 & 0 & 0 & 0 & -\mathbf{y}_i^u \mathbf{x}_i^u & -\mathbf{y}_i^u \mathbf{x}_i^v \\ 0 & 0 & 0 & \mathbf{x}_i^u & \mathbf{x}_i^v & 1 & -\mathbf{y}_i^v \mathbf{x}_i^u & -\mathbf{y}_i^v \mathbf{x}_i^v \end{bmatrix}, \tag{11}$$

we have the following equation:

$$\tilde{\mathbf{W}}_i \tilde{\mathbf{h}} = \mathbf{y}_i. \tag{12}$$

Thus, the Gaussian field criterion becomes:

$$G(\tilde{\mathbf{h}}) = - \sum_{i=1}^N e^{-\frac{\|\mathbf{y}_i - \tilde{\mathbf{W}}_i \tilde{\mathbf{h}}\|^2}{\sigma^2}}. \tag{13}$$

The above objective function is always continuously differentiable with respect to the homography parameter $\tilde{\mathbf{h}}$, and its derivative has the following form:

$$\frac{\partial G(\tilde{\mathbf{h}})}{\partial \tilde{\mathbf{h}}} = \sum_{i=1}^N \frac{2(\tilde{\mathbf{W}}_i^T \tilde{\mathbf{W}}_i \tilde{\mathbf{h}} - \tilde{\mathbf{W}}_i^T \mathbf{y}_i)}{\sigma^2} e^{-\frac{\|\mathbf{y}_i - \tilde{\mathbf{W}}_i \tilde{\mathbf{h}}\|^2}{\sigma^2}}. \tag{14}$$

By using the derivative in Eq. (14), we adopt gradient-based numerical optimization techniques such as the quasi-Newton method to solve the homography parameter $\tilde{\mathbf{h}}$. In addition, we apply the deterministic annealing technique [54] on the range parameter σ to improve the convergence, where σ is initialized with a large value and gradually reduced by $\sigma \rightarrow \gamma\sigma$ with γ being the annealing rate.

The correctness of a putative match could be determined by checking whether it is consistent with \mathbf{H} . We summarize the proposed Gaussian field criterion (GFC) method for linear matching in Algorithm 1.

Algorithm 1: The Gaussian field criterion for linear matching

Input: Putative match set S , annealing rate γ

Output: Inlier set \mathcal{I}

- 1 Calculate the coefficient matrix $\tilde{\mathbf{W}}_i$;
 - 2 Initialize parameter σ ;
 - 3 Deterministic annealing:
 - 4 Use gradient (14) to optimize objective function (13);
 - 5 Update homography $\tilde{\mathbf{h}} \leftarrow \arg \min_{\tilde{\mathbf{h}}} G(\tilde{\mathbf{h}})$;
 - 6 Anneal $\sigma = \gamma\sigma$;
 - 7 The inlier set \mathcal{I} is determined based on the consistency between the putative match set S and homography \mathbf{H} .
-

3.4 Matching with nonlinear transformation model

We next consider the matching problem undergoing nonlinear transformation. We utilize a general nonparametric model to handle complex (e.g., non-rigid) transformations. Specifically, the transformation \mathbf{f} is modeled by requiring it to lie within a functional space \mathcal{H} , namely reproducing kernel Hilbert space (RKHS) [55]. We define \mathcal{H} by a diagonal decomposable kernel Γ : $\Gamma(\mathbf{x}_i, \mathbf{x}_j) = e^{-\beta\|\mathbf{x}_i - \mathbf{x}_j\|^2} \cdot \mathbf{I}$; then, \mathbf{f} has the following form:

$$\mathbf{f}(\mathbf{x}) = \sum_{i=1}^N \Gamma(\mathbf{x}, \mathbf{x}_i) \mathbf{c}_i \tag{15}$$

with $\{\mathbf{c}_i\}_{i=1}^N$ being the coefficients.

The non-rigid transformation will lead to ill-posedness in Eq. (4), as the solution of \mathbf{f} is not unique. To ensure well-posedness, we introduce a regularization term to control the complexity of \mathbf{f} and the Gaussian field criterion for nonlinear matching then becomes:

$$G(\mathbf{f}) = - \sum_{i=1}^N e^{-\frac{d^2(\mathbf{f}(\mathbf{x}_i), \mathbf{y}_i)}{\sigma^2}} + \lambda \|\mathbf{f}\|_{\mathcal{H}}^2, \tag{16}$$

where $\lambda > 0$ controls the trade-off between the first data fidelity term and the second regularization term and $\|\cdot\|_{\mathcal{H}}$ is a functional norm which can be defined by an inner product, e.g., $\|\mathbf{f}\|_{\mathcal{H}}^2 = \langle \mathbf{f}, \mathbf{f} \rangle_{\mathcal{H}}$.

To optimize the objective function in Eq. (16), we rewrite it in the matrix form:

$$G(\mathbf{C}) = - \sum_{i=1}^N e^{-\frac{\|\mathbf{y}_i^T - \Gamma_{i, \cdot} \mathbf{C}\|^2}{\sigma^2}} + \lambda \cdot \text{tr}(\mathbf{C}^T \mathbf{\Gamma} \mathbf{C}), \tag{17}$$

where $\mathbf{\Gamma} \in \mathbb{R}^{N \times N}$ with $\Gamma_{ij} = e^{-\beta\|\mathbf{x}_i - \mathbf{x}_j\|^2}$ is the so-called Gram matrix, $\mathbf{C} = (\mathbf{c}_1, \dots, \mathbf{c}_N)^T$, and $\text{tr}(\cdot)$ is the trace.

The objective function in Eq. (17) is always continuously differentiable with respect to \mathbf{C} , and its derivative has the following form:

$$\frac{\partial G(\mathbf{C})}{\partial \mathbf{C}} = - \sum_{i=1}^N \frac{2\mathbf{\Gamma}_{i,\cdot}^T (\mathbf{\Gamma}_{i,\cdot} \mathbf{C} - \mathbf{y}_i^T)}{\sigma^2} e^{-\frac{\|\mathbf{y}_i^T - \mathbf{\Gamma}_{i,\cdot} \mathbf{C}\|^2}{\sigma^2}} + 2\lambda \mathbf{\Gamma} \mathbf{C}. \tag{18}$$

Similar to the linear matching algorithm, by using the derivative in Eq. (18) and some numerical optimization techniques together with the deterministic annealing technique, we can solve the transformation parameter \mathbf{C} .

Fast Implementation. The Gram matrix $\mathbf{\Gamma}$ is of size $N \times N$, and hence, it will typically cost $O(N^3)$ to solve \mathbf{C} . This will pose significant computational burden when dealing with real-time or large-scale problems. To address this issue, we apply a sparse approximation to the transformation \mathbf{f} based on the subset of regressors method [56, 57]. Specifically, we randomly pick a subset of M points $\{\tilde{\mathbf{x}}_i\}_{i=1}^M$ and only let them have nonzero coefficients in Eq. (15); thus, \mathbf{f} becomes:

$$\mathbf{f}^s(\mathbf{x}) = \sum_{i=1}^M \mathbf{\Gamma}(\mathbf{x}, \tilde{\mathbf{x}}_i) \mathbf{c}_i^s. \tag{19}$$

It has been shown in [56] that selecting subset randomly performs no worse than those more time-consuming and sophisticated methods. Accordingly, the derivative becomes:

$$\frac{\partial G(\mathbf{C}^s)}{\partial \mathbf{C}^s} = - \sum_{i=1}^N \frac{2\mathbf{U}_{i,\cdot}^T (\mathbf{U}_{i,\cdot} \mathbf{C}^s - \mathbf{y}_i^T)}{\sigma^2} e^{-\frac{\|\mathbf{y}_i^T - \mathbf{U}_{i,\cdot} \mathbf{C}^s\|^2}{\sigma^2}} + 2\lambda \mathbf{\Gamma}^s \mathbf{C}^s, \tag{20}$$

where the definitions of $\mathbf{\Gamma}^s \in \mathbb{R}^{M \times M}$ and $\mathbf{C}^s \in \mathbb{R}^{M \times 2}$ are similar to $\mathbf{\Gamma}$ and \mathbf{C} , $\mathbf{U} \in \mathbb{R}^{N \times M}$ with $\mathbf{U}_{ij} = e^{-\beta \|\mathbf{x}_i - \tilde{\mathbf{x}}_j\|^2}$.

The correctness of a putative match could be determined by checking whether it is consistent with \mathbf{f} . We summarize the proposed Gaussian field criterion (GFC) method for nonlinear matching in Algorithm 2.

Algorithm 2: The Gaussian field criterion for nonlinear matching

Input: Putative match set S , parameters $\beta, \lambda, \gamma, M$

Output: Inlier set \mathcal{I}

- 1 Calculate Gram matrix $\mathbf{\Gamma}^s$ and matrix \mathbf{U} ;
 - 2 Initialize coefficient matrix \mathbf{C}^s and parameter σ ;
 - 3 Deterministic annealing;
 - 4 Use gradient (20) to optimize objective function (16);
 - 5 Update coefficients $\mathbf{C}^s \leftarrow \arg \min_{\mathbf{C}^s} G(\mathbf{C}^s)$;
 - 6 Anneal $\sigma = \gamma \sigma$;
 - 7 The inlier set \mathcal{I} is determined based on the consistency between the putative match set S and transformation \mathbf{f} .
-

3.5 Computational complexity

The major computational burden is to compute the derivative in Eq. (14) or (18) and solve the transformation parameter using a numerical technique such as the quasi-Newton method. For linear matching, they cost both about $O(N)$ and hence, the total time complexity is about $O(N)$. The space complexity is about $O(N)$ due to the requirement of storing the N coefficient matrix \mathbf{W}_i .

For the nonlinear matching in Algorithm 2, without using the sparse approximation, the calculation of the Gram matrix $\mathbf{\Gamma}$ in Line 1 costs $O(N^2)$. The computation of derivative in Eq. (18) and optimization of transformation with quasi-Newton method cost, respectively, about $O(N^2)$ and $O(N^3)$, and hence, the total time complexity is about $O(N^3)$. The space complexity is about $O(N^2)$ due to the requirement of storing the Gram matrix $\mathbf{\Gamma}$. By using the sparse approximation, the calculations of matrices $\mathbf{\Gamma}^s$ and \mathbf{U} in Line 1 cost, respectively, $O(M^2)$ and $O(MN)$, while the time complexities of computing the derivative and solving \mathbf{C} , respectively, reduce to about $O(MN)$ and $O(M^3)$. Therefore, the total time complexity reduces to about $O(MN + M^3)$. As M is a constant and $M \ll N$, we can simply write the time complexity as $O(N)$. Clearly, the space complexity is also reduced to $O(N)$. In practice, the sparse approximation is always encouraged to achieve fast implementations, especially in real-time matching problems.

In conclusion, our Gaussian field criterions for linear and nonlinear matching both have linear time complexity and linear space complexity, which is significant for handling real-time problems or matching high-resolution images with plenty of feature points.

3.6 Implementation details

The performance of feature matching typically depends on the coordinate system which is used to express the feature points; here, we use data normalization as that in [50] to control this. More specifically, we apply a linear scaling to each point set, so that it has zero mean and unit variance.

For the linear matching problem, there is only one parameter in our GFC such as γ , the deterministic annealing rate used for improving the convergence of transformation estimation. We fix it to 0.93 throughout the experiments. For the nonlinear matching problem, three additional parameters are involved such as β, λ , and M . Parameters β and λ both control the smoothness of the transformation \mathbf{f} , where the former determines the width of interaction range between feature points through the transformation and the latter constrains the complexity of the transformation. Parameter M determines the number of

basis used for sparse approximation to construct the transformation. We fix them throughout the experiments as follows: $\beta = 0.01$, $\lambda = 0.1$, and $M = 30$.

4 Experimental results

In this section, we test our GFC for feature matching on various real image data including both natural images and remote sensing images and compare it with five other state-of-the-art feature matching methods including RANSAC [41], ICF [44], GS [45], VFC [14], LPM [58], and MR [47]. All the experiments are conducted on a laptop with 2.4 GHz Intel Core CPU, 8 GB memory, and MATLAB code.

4.1 Datasets and settings

To perform a comprehensive evaluation of our method, we conduct experiments on both natural and remote sensing image datasets. In the following, we first describe the test data and then introduce the evaluation criteria.

- *Natural image dataset* The publicly available VGG dataset [59] is adopted for linear matching. This dataset contains eight groups of images, and each group consists of five image pairs, leading to 40 image pairs in total. The image pairs suffer from different geometric and photometric transformations including viewpoint change, scale change, rotation, light change, image blur, and JPEG compression. In addition, these image pairs are related always by homography due to that they are either of planar scenes or acquired in a fixed camera position. The ground truth homographies are provided in the dataset. Some example image pairs are shown in Fig. 3. To make the dataset more

challenging, we construct two putative sets from each image pair by setting different distance ratio thresholds to measure the similarity between SIFT descriptors. For nonlinear matching, we collect several typical image pairs with relatively complex deformations, including wide baseline pairs and image pairs involving non-rigid deformations.

- *Remote sensing image dataset* For linear matching, we collect several color–infrared aerial photograph image pairs and SPOT image pairs, which involve translation, rotation, scaling, and so on. For nonlinear matching, we collect a set of 50 image pairs including SPOT images, SAR images, panchromatic aerial photographs, as well as UAV images. These image pairs usually involve ground relief variations and imaging viewpoint changes, where the image transformations do not exactly satisfy a linear model such as homography. Therefore, a more complex nonlinear model is necessary to generate accurate matching results.

We use the open-source VLFeat toolbox [60] to determine the putative SIFT matches. The putative matches are then established based on the similarity between SIFT descriptors. The matching performance is characterized by precision and recall, where the precision is defined as the ratio between the identified correct match number and the whole preserved match number, and the recall is defined as the ratio between the identified correct match number and the whole correct match number.

To establish the ground truth feature matches in each image pair, for linear matching, we first acquire the ground truth homography and then an overlap error is used to determine the match correctness as that in [14]. The ground truth homography in the remote sensing image dataset is calculated from manually selected correspondences together with least squares. For nonlinear matching, we manually check the match correctness of each putative match for all image pairs.

4.2 Results on natural image pairs

We first evaluate the matching performance of our GFC with linear model on the VGG dataset [59]. The precision–recall (p–r) statistics of different methods are reported in the middle of Fig. 4. Each scatter point corresponds to a p–r pair of a certain method on a certain image pair. The average match number in the putative sets is about 773.9. From the results, we see that all the six methods work quite well on this dataset. In general, our GFC has a better trade-off compared to the other methods and the p–r pairs of our method almost concentrate on the top right corner. More specifically, the average p–r pair achieves about (97.81, 97.57%). We also report the run-time statistics of

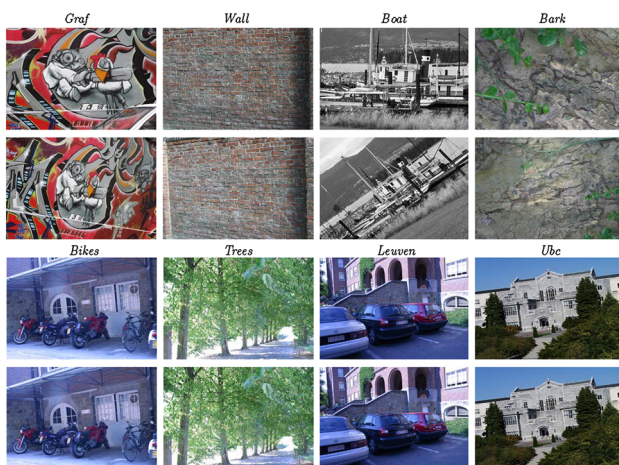


Fig. 3 Example image pairs in the VGG dataset [59]

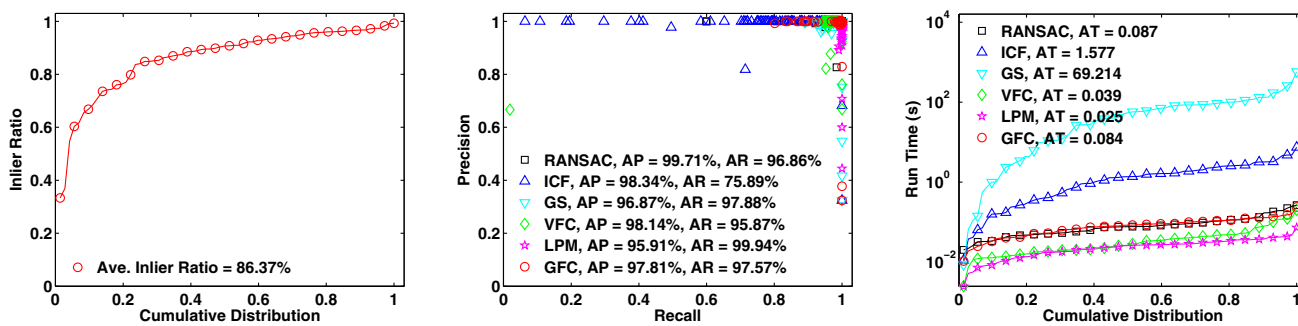


Fig. 4 Statistics of initial inlier ratio (left), precision–recall pair (middle), and run-time (right) of different methods on the VGG dataset [59]

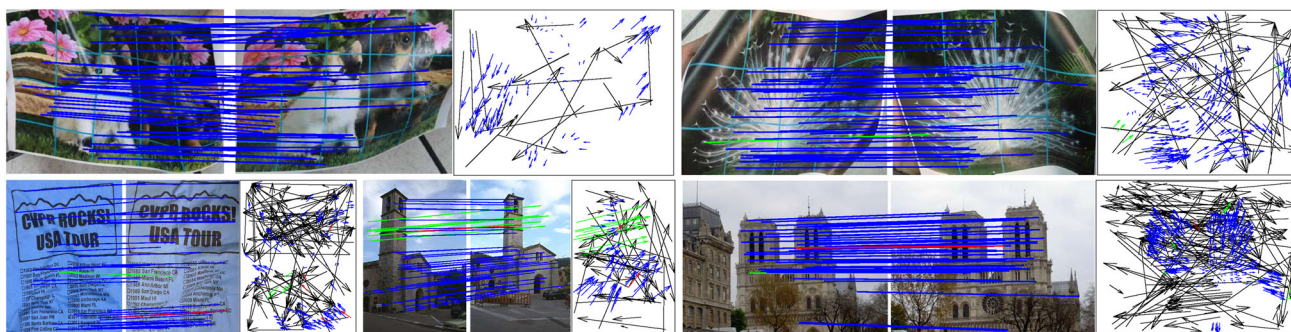


Fig. 5 Qualitative matching results of our GFC with linear model on *NotreDame*, *Church*, *DogCat*, *Peacock*, and *T-shirt*, respectively. The inlier ratios of 5 image pairs are 75.27, 54.76, 81.74, 83.63, and 49.64%. The head and tail of each arrow in the motion field correspond to the positions of feature points in two images (blue =

true positive, black = true negative, green = false negative, and red = false positive). For visibility, in the image pairs, at most 50 randomly selected matches are presented, and the true negatives are not shown. Best viewed in color

Table 1 Quantitative comparison of different methods on the image pairs in Fig. 5

	<i>NotreDame</i>	<i>Church</i>	<i>DogCat</i>	<i>Peacock</i>	<i>T-shirt</i>
ICF [44]	(100.0, 77.86)	(85.71, 78.26)	(82.30, 98.94)	(99.18, 86.48)	(94.90, 67.39)
GS [45]	(100.0, 87.30)	(92.86, 94.20)	(97.89, 98.94)	(100.0, 86.12)	(97.46, 83.33)
VFC [14]	(100.0, 93.83)	(98.24, 81.16)	(100.0, 98.94)	(99.63, 96.44)	(96.97, 92.75)
LPM [58]	(75.40, 94.56)	(54.46, 88.41)	(82.00, 87.23)	(82.87, 94.66)	(52.07, 63.77)
MR [47]	(99.60, 90.56)	(95.00, 82.61)	(100.0, 100.0)	(99.64, 98.58)	(97.71, 92.75)
GFC	(99.45, 98.37)	(96.77, 86.96)	(100.0, 100.0)	(99.63, 98.22)	(95.68, 96.38)

The pairs in the table are precision–recall pairs (%)

different methods, as shown in the right figure of Fig. 4. We see that LPM is the most efficient one and our GFC ranks middle with an average run-time of about 84 milliseconds.

We next provide some intuitionistic matching results of our GFC with nonlinear model on several typical image pairs with relatively complex deformations, as shown in Fig. 5. The first two pairs are wide baseline images, and the rest three pairs involve non-rigid deformations. For each group of results, the left image pair schematically shows the matching result, while the right field presents the matching correctness of each putative match. From the

results, we again see that our GFC is robust and accurate and very few putative matches are misjudged.

In addition, we also provide quantitative comparison of different methods on these image pairs, as shown in Table 1. As RANSAC relies on a parametric model which is not suitable for non-rigid matching, we do not report its results in this table. From the results, we see that ICF and GS usually cannot obtain good precision and recall simultaneously, while LPM cannot work well in case of low initial inlier percentage. VFC and MR in general can produce satisfying results. However, our GFC clearly has the best precision–recall trade-off on all the five image pairs.

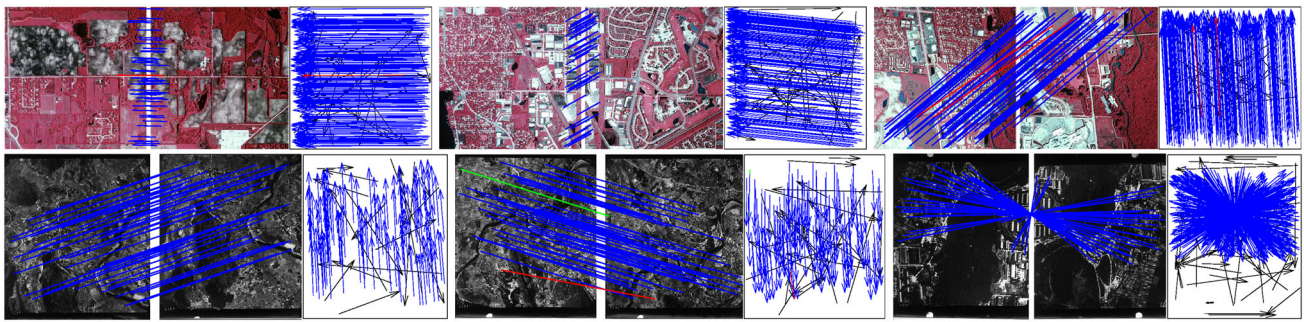


Fig. 6 Qualitative matching results of our GFC with linear model on six typical remote sensing image pairs involving linear transformation. The inlier ratios of 6 image pairs are 90.15, 84.14, 90.12, 87.97, 85.31, and 89.98%. The head and tail of each arrow in the motion field correspond to the positions of feature points in two images (blue

= true positive, black = true negative, green = false negative and red = false positive). For visibility, in the image pairs, at most 50 randomly selected matches are presented and the true negatives are not shown. Best viewed in color

Table 2 Quantitative comparison of different methods on the image pairs in Fig. 6

	<i>Image1</i>	<i>Image2</i>	<i>Image3</i>	<i>Image4</i>	<i>Image5</i>	<i>Image6</i>
RANSAC [41]	(100.0, 100.0)	(100.0, 100.0)	(100.0, 100.0)	(100.0, 99.15)	(100.0, 100.0)	(100.0, 98.70)
ICF [44]	(99.66, 99.32)	(100.0, 82.12)	(100.0, 100.0)	(96.69, 100.0)	(88.30, 100.0)	(100.0, 99.74)
GS [45]	(100.0, 100.0)	(100.0, 100.0)	(100.0, 100.0)	(88.64, 100.0)	(86.29, 100.0)	(100.0, 99.74)
VFC [14]	(100.0, 99.66)	(100.0, 100.0)	(100.0, 100.0)	(100.0, 94.87)	(100.0, 98.01)	(89.98, 100.0)
LPM [58]	(100.0, 100.0)	(99.32, 100.0)	(100.0, 100.0)	(100.0, 99.15)	(100.0, 98.01)	(100.0, 99.74)
MR [47]	(100.0, 99.66)	(100.0, 100.0)	(100.0, 100.0)	(100.0, 94.87)	(100.0, 98.01)	(89.98, 100.0)
GFC	(99.66, 100.0)	(100.0, 100.0)	(99.32, 100.0)	(100.0, 100.0)	(99.34, 99.34)	(98.97, 100.0)

The pairs in the table are precision–recall pairs (%)

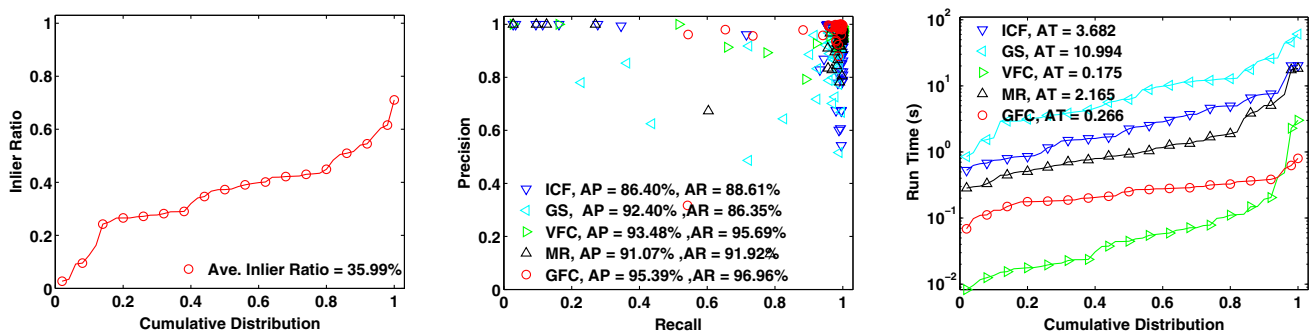


Fig. 7 Statistics of initial inlier ratio (left), precision–recall pair (middle), and run-time (right) of different methods on the remote sensing dataset

4.3 Results on remote sensing image pairs

Now we test our GFC on the remote sensing image datasets. First, we test the linear model on several typical image pairs involving only linear transformation, as shown in Fig. 6. From the results, we see that our GFC is able to always produce satisfying matching performance and almost all the inliers and outliers are accurately identified. We also provide the results of the other six comparison methods in Table 2. Again, all the seven methods' results are quite satisfying, which is mainly due to the simplicity

of linear matching and the large initial inlier ratios in the image pairs.

We further report the matching performance of our GFC with nonlinear model on a remote sensing image dataset with 50 image pairs. The statistic results are shown in Fig. 7. To make the dataset more challenging, we use the nearest neighbor strategy to construct the putative set in each image pair. For example, for each feature point in one image, we search its nearest neighbor in the other image in descriptor space under Euclidean distance, and then these two points are considered as a putative match. This

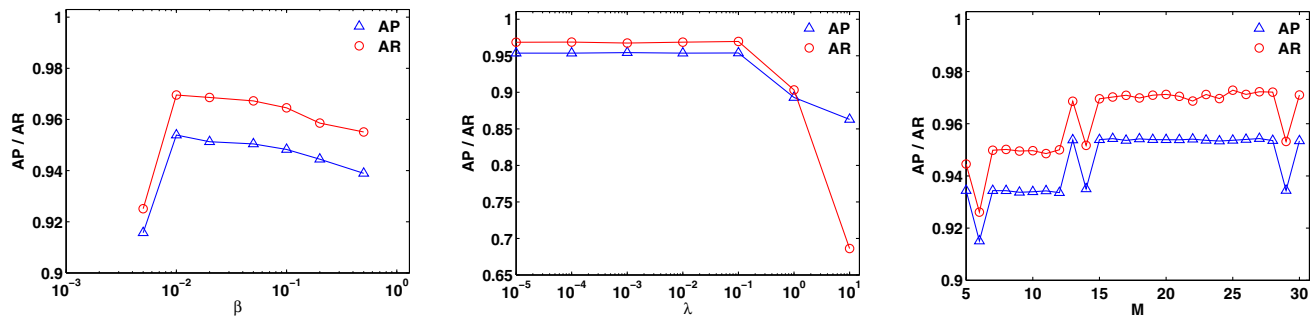


Fig. 8 Average precisions and recalls on the remote sensing dataset with respect to different values of β (left), λ (middle), and M (right)

procedure results in very low initial inlier ratios in the putative sets, as shown in the left figure of Fig. 7. The average initial inlier ratio is merely 35.99%. Here we do not provide the results of LPM either, as it cannot obtain satisfying results in case of such low inlier ratio. The performance comparison of the rest five methods is shown in the middle figure of Fig. 7. Clearly, our GFC has the best results, in terms of both precision and recall. That is to say, our GFC is much more robust in case of large outliers. We also report the run-time statistics of different methods in the right figure of Fig. 7. Our GFC again ranks middle with an average run-time of about 0.678 s. Moreover, we also have tested the performance of our GFC without sparse approximation on this dataset and obtained an average p-r pair of about (95.05, 96.82%). However, the average run-time exceeds 2000 s. This demonstrates that the sparse approximation can significantly reduce the computational complexity without sacrificing the matching accuracy.

Finally, we test the influence of parameter settings of our GFC on the remote sensing image dataset. There are mainly three parameters such as β , λ , and M . We vary them and compute the average p-r pairs on the whole dataset. The statistic results are shown in Fig. 8. In each subfigure, we vary one parameter and fix the other parameter to their “optimal values.”² From the results, we see that the average precision and recall both achieve their largest values at $\beta = 0.01$ and $\lambda = 0.1$. For parameter M , the average precision and recall do not change too much when $M \geq 15$. Nevertheless, $M = 15$ is more preferable for computational efficiency.

5 Conclusion

Within this paper, we introduce a robust estimator named Gaussian field criterion (GFC) for image feature matching. A key characteristic of our GFC is that it inherently has robustness to outliers and hence can efficiently remove

false matches from a set of putative matches. The image transformation is modeled by both linear and nonlinear functions, and a sparse approximation is applied to the nonlinear case to significantly reduce the computational complexity. The qualitative and quantitative experiments on publicly available datasets, including both natural and remote sensing image pairs, demonstrate that our GFC is quite efficient for mismatch removal. More specifically, it is able to produce superior matching performance over several state-of-the-arts, especially when the image scenes suffer from complex non-rigid deformations or lots of outliers. This will be of great benefit to registering remote sensing images with low overlaps, repeated patterns, local distortions or low quality.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant Nos. 61773295 and 61503288.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest

References

- Jiang, J., Chen, C., Ma, J., Wang, Z., Wang, Z., Hu, R.: Srlsp: A face image super-resolution algorithm using smooth regression with local structure prior. *IEEE Trans. Multimed.* **19**(1), 27–40 (2017)
- Ma, J., Chen, C., Li, C., Huang, J.: Infrared and visible image fusion via gradient transfer and total variation minimization. *Inf. Fusion* **31**, 100–109 (2016)
- Li, Y., Huang, X., Liu, H.: Unsupervised deep feature learning for urban village detection from high-resolution remote sensing images. *Photogramm. Eng. Remote Sens.* **83**(8), 567–579 (2017)
- Li, Y., Tao, C., Tan, Y., Shang, K., Tian, J.: Unsupervised multilayer feature learning for satellite image scene classification. *IEEE Geosci. Remote Sens. Lett.* **13**(2), 157–161 (2016)
- Ma, J., Ma, Y., Li, C.: Infrared and visible image fusion methods and applications: a survey. *Inf. Fusion.* **45**, 153–178 (2018)
- Liu, T., Liu, H., Chen, Z., Lesgold, A.M.: Fast blind instrument function estimation method for industrial infrared spectrometers. *IEEE Trans. Ind. Inf.* (2018). <https://doi.org/10.1109/TII.2018.2794449>

² In this paper, we consider the optimal values as: $\beta = 0.01$, $\lambda = 0.1$, and $M = 30$.

7. Liu, Y., Chen, X., Peng, H., Wang, Z.: Multi-focus image fusion with a deep convolutional neural network. *Inf. Fusion* **36**, 191–207 (2017)
8. Gao, C., Wang, L., Xiao, Y., Zhao, Q., Meng, D.: Infrared small-dim target detection based on Markov random field guided noise modeling. *Pattern Recognit.* **76**, 463–475 (2018)
9. Liu, Y., Chen, X., Wang, Z., Wang, Z.J., Ward, R.K., Wang, X.: Deep learning for pixel-level image fusion: recent advances and future prospects. *Inf. Fusion* **42**, 158–173 (2018)
10. Ma, J., Zhou, H., Zhao, J., Gao, Y., Jiang, J., Tian, J.: Robust feature matching for remote sensing image registration via locally linear transforming. *IEEE Trans. Geosci. Remote Sens.* **53**(12), 6469–6481 (2015)
11. Wong, A., Clausi, D.A.: ARRSI: automatic registration of remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **45**(5), 1483–1493 (2007)
12. Zitová, B., Flusser, J.: Image registration methods: a survey. *Image Vis. Comput.* **21**, 977–1000 (2003)
13. Lowe, D.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
14. Ma, J., Zhao, J., Tian, J., Yuille, A.L., Tu, Z.: Robust point matching via vector field consensus. *IEEE Trans. Image Process.* **23**(4), 1706–1721 (2014)
15. Jiang, J., Hu, R., Wang, Z., Han, Z., Ma, J.: Facial image hallucination through coupled-layer neighbor embedding. *IEEE Trans. Circuits Syst. Video Technol.* **26**(9), 1674–1684 (2016)
16. Jiang, J., Ma, J., Chen, C., Jiang, X., Wang, Z.: Noise robust face image super-resolution through smooth sparse representation. *IEEE Trans. Cybern.* **47**(11), 3991–4002 (2017)
17. Maier, J., Humenberger, M., Murschitz, M., Zendel, O., Vincze, M.: Guided matching based on statistical optical flow for fast and robust correspondence analysis. In: *Proceedings of European Conference on Computer Vision*. pp. 101–117 (2016)
18. Gao, Y., Ma, J., Yuille, A.L.: Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples. *IEEE Trans. Image Process.* **26**(5), 2545–2560 (2017)
19. Liu, Y., De Dominicis, L., Wei, B., Chen, L., Martin, R.R.: Regularization based iterative point match weighting for accurate rigid transformation estimation. *IEEE Trans. Vis. Comput. Graph.* **21**(9), 1058–1071 (2015)
20. Yang, K., Pan, A., Yang, Y., Zhang, S., Ong, S.H., Tang, H.: Remote sensing image registration using multiple image features. *Remote Sens.* **9**(6), 581 (2017)
21. Guo, X., Li, Y., Ling, H.: Lime: low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **26**(2), 982–993 (2017)
22. Ma, J., Jiang, J., Liu, C., Li, Y.: Feature guided Gaussian mixture model with semi-supervised em and local geometric constraint for retinal image registration. *Inf. Sci.* **417**, 128–142 (2017)
23. Wang, G., Wang, Z., Chen, Y., Zhao, W.: Robust point matching method for multimodal retinal image registration. *Biomed. Signal Process. Control* **19**, 68–76 (2015)
24. Li, J., Hu, Q., Ai, M.: Robust feature matching for remote sensing image registration based on I_q -estimator. *IEEE Geosci. Remote Sens. Lett.* **13**(12), 1989–1993 (2016)
25. Li, J., Hu, Q., Ai, M., Zhong, R.: Robust feature matching via support-line voting and affine-invariant ratios. *ISPRS J. Photogramm. Remote Sens.* **132**, 61–76 (2017)
26. Li, Y., Zhang, Y., Huang, X., Zhu, H., Ma, J.: Large-scale remote sensing image retrieval by deep hashing neural networks. *IEEE Trans. Geosci. Remote Sens.* **52**(2), 950–965 (2018)
27. Shi, X., Jiang, J.: Automatic registration method for optical remote sensing images with large background variations using line segments. *Remote Sens.* **8**(5), 426 (2016)
28. Wei, Z., Han, Y., Li, M., Yang, K., Yang, Y., Luo, Y., Ong, S.-H.: A small UAV based multi-temporal image registration for dynamic agricultural terrace monitoring. *Remote Sens.* **9**(9), 904 (2017)
29. Gonzalez, R.C., Wintz P.: *Digital image processing*. New York, NY, USA: Addison-Wesley (1987)
30. Reddy, B.S., Chatterji, B.N.: An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Trans. Image Process.* **5**(8), 1266–1271 (1996)
31. Rangarajan, A., Chui, H., Duncan, J.S.: Rigid point feature registration using mutual information. *Med. Image Anal.* **3**(4), 425–440 (1999)
32. Le Moigne, J., Campbell, W.J., Crompt, R.F.: An automated parallel image registration technique based on the correlation of wavelet features. *IEEE Trans. Geosci. Remote Sens.* **40**(8), 1849–1864 (2002)
33. Chen, Q.-S., Defrise, M., Deconinck, F.: Symmetric phase-only matched filtering of Fourier–Mellin transforms for image registration and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**(12), 1156–1168 (1994)
34. Uss, M.L., Vozel, B., Lukin, V.V., Chehdi, K.: Multimodal remote sensing image registration with accuracy estimation at local and global scales. *IEEE Trans. Geosci. Remote Sens.* **54**(11), 6587–6605 (2016)
35. Guo, X., Cao, X.: Good match exploration using triangle constraint. *Pattern Recognit. Lett.* **33**(7), 872–881 (2012)
36. Pele, O., Werman, M.: A linear time histogram metric for improved SIFT matching. In: *Proceedings of European Conference on Computer Vision*. pp. 495–508 (2008)
37. Li, Q., Wang, G., Liu, J., Chen, S.: Robust scale-invariant feature matching for remote sensing image registration. *IEEE Geosci. Remote Sens. Lett.* **6**(2), 287–291 (2009)
38. Hu, Y.-T., Lin, Y.-Y., Chen, H.-Y., Hsu, K.-J., Chen, B.-Y.: Matching images with multiple descriptors: an unsupervised approach for locally adaptive descriptor selection. *IEEE Trans. Image Process.* **24**(12), 5995–6010 (2015)
39. Wang, C., Wang, L., Liu, L.: Progressive mode-seeking on graphs for sparse feature matching. In: *Proceedings of European Conference on Computer Vision*. pp. 788–802 (2014)
40. Cho, M., Lee, K.M.: Progressive graph matching: making a move of graphs via probabilistic voting. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 398–405, (2012)
41. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. ACM* **24**(6), 381–395 (1981)
42. Torr, P.H.S., Zisserman, A.: MLESAC: a new robust estimator with application to estimating image geometry. *Comput. Vis. Image Understand.* **78**(1), 138–156 (2000)
43. Chum, O., Matas, J.: Matching with PROSAC—progressive sample consensus. In: *Proceedings of IEEE Conference on Computer Vision Pattern Recognition*. pp. 220–226 (2005)
44. Li, X., Hu, Z.: Rejecting mismatches by correspondence function. *Int. J. Comput. Vis.* **89**(1), 1–17 (2010)
45. Liu, H., Yan, S.: Common visual pattern discovery via spatially coherent correspondence. In: *Proceedings of IEEE Conference on Computer Vision Pattern Recognition*. pp. 1609–1616 (2010)
46. Yao, G., Cui, J., Deng, K., Zhang, L.: Robust Harris corner matching based on the quasi-homography transform and self-adaptive window for wide-baseline stereo images. *IEEE Trans. Geosci. Remote Sens.* **56**(1), 559–574 (2018)
47. Yu, Z., Zhou, H., Li, C.: Fast non-rigid image feature matching for agricultural UAV via probabilistic inference with regularization techniques. *Comput. Electron. Agric.* **143**, 79–89 (2017)
48. Ma, J., Zhao, J., Jiang, J., Zhou, H.: Non-rigid point set registration with robust transformation estimation under manifold

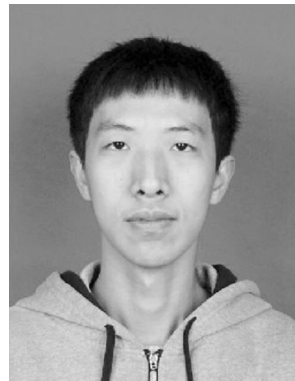
- regularization. In: Proceedings on AAAI Conference of Artificial Intelligence. pp. 4218–4224 (2017)
49. Boughorbel, F., Koschan, A., Abidi, B., Abidi, M.: Gaussian fields: a new criterion for 3D rigid registration. *Pattern Recognit.* **37**(7), 1567–1571 (2004)
 50. Ma, J., Zhao, J., Ma, Y., Tian, J.: Non-rigid visible and infrared face registration via regularized Gaussian fields criterion. *Pattern Recognit.* **48**(3), 772–784 (2015)
 51. Wang, G., Zhou, Q., Chen, Y.: Robust non-rigid point set registration using spatially constrained Gaussian fields. *IEEE Trans. Image Process.* **26**(4), 1759–1769 (2017)
 52. Wang, G., Chen, Y., Zheng, X.: Gaussian field consensus: a robust nonparametric matching method for outlier rejection. *Pattern Recognit.* **74**, 305–316 (2018)
 53. Greengard, L., Strain, J.: The fast Gauss transform. *SIAM J. Sci. Stat. Comput.* **12**(1), 79–94 (1991)
 54. Yuille, A.L.: Generalized deformable models, statistical physics, and matching problems. *Neural Comput.* **2**(1), 1–24 (1990)
 55. Micchelli, C.A., Pontil, M.: On learning vector-valued functions. *Neural Comput.* **17**(1), 177–204 (2005)
 56. Ma, J., Zhao, J., Tian, J., Bai, X., Tu, Z.: Regularized vector field learning with sparse approximation for mismatch removal. *Pattern Recognit.* **46**(12), 3519–3532 (2013)
 57. Liu, H., Liu, S., Huang, T., Zhang, Z., Hu, Y., Zhang, T.: Infrared spectrum blind deconvolution algorithm via learned dictionaries and sparse representation. *Appl. Opt.* **55**(10), 2813–2818 (2016)
 58. Ma, J., Zhao, J., Guo, H., Jiang, J., Zhou, H., Gao, Y.: Locality preserving matching. In: Proceedings of International Joint Conference on Artificial Intelligence. pp. 4492–4498 (2017)
 59. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., van Gool, L.: A comparison of affine region detectors. *Int. J. Comput. Vis.* **65**(1), 43–72 (2005)
 60. Vedaldi, A., Fulkerson, B.: VLFeat—an open and portable library of computer vision algorithms. In: Proceedings of the ACM International Conference on Multimedia. pp. 1469–1472 (2010)



Qing Ma received the B.S. degree from the School of Materials and Metallurgy, Wuhan University of Science and Technology, Wuhan, China, in 2015. He is currently a Master student with the National Engineering Research Center for E-Learning, Central China Normal University. His research interests are in the areas of computer vision and pattern recognition.



Xu Du received the Ph.D. degree in electronics and information system from Huazhong University of Science and Technology, Wuhan, China, in 2002. He is currently a professor with National Engineering Research Center for E-learning, Central China Normal University, and his current research interests include machine learning, cloud and mobile learning, multimedia network technology and applications.



Jiahao Wang is currently an undergraduate with the Electronic Information School, Wuhan University, majoring in Electronic Engineering. His research interests are in the areas of computer vision and pattern recognition.



Yong Ma received the Degree from the Department of Automatic Control, Beijing Institute of Technology, Beijing, China, in 1997 and the Ph.D. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2003. From 2004 to 2006, he was a Lecturer with the University of the West of England, Bristol, the UK. From 2006 to 2014, he was with the Wuhan National Laboratory for Optoelectronics, HUST, where he was a Professor of Electronics. He is currently a Professor with the Electronic Information School, Wuhan University, Wuhan. His current research interests include signals and systems, remote sensing of the lidar and infrared, infrared image processing, pattern recognition, and interface circuits to sensors and actuators.



Jiayi Ma received the B.S. degree from the Department of Mathematics and the Ph.D. degree from the School of Automation, Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. From 2012 to 2013, he was an Exchange Student with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA, the USA. He is currently an Associate Professor with the Electronic Information School, Wuhan University, Wuhan, where he holds a postdoctoral

position from 2014 to 2015. Prof. Ma is an Associate Editor of IEEE Access. He has authored or co-authored over 80 refereed journal and conference papers. His current research interests include computer vision, machine learning, and pattern recognition.