ISPRS Journal of Photogrammetry and Remote Sensing 183 (2022) 196–209

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Robust feature matching via neighborhood manifold representation consensus

Jiayi Ma [a], Zizhuo Li [a], Kaining Zhang [a], Zhenfeng Shao [b,*], Guobao Xiao [c]

[a] Electronic Information School, Wuhan University, Wuhan 430072, China
[b] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[c] College of Computer and Control Engineering, Minjiang University, Fuzhou 350108, China

ARTICLE INFO

Keywords:
Feature matching
Local neighborhood structure
Manifold representation
Image registration
Outlier

ABSTRACT

Feature matching, which aims at seeking dependable correspondences between two sets of features, is of considerable significance to various vision-based tasks. This paper attempts to eliminate false correspondences from given tentative correspondences created on the basis of descriptor similarity. A simple yet efficient approach named *neighborhood manifold representation consensus* (NMRC) for robust feature matching is presented considering the stable neighborhood topologies of the potential true matches. The core principle of the proposed method is to preserve the local neighborhood structures between two feature points in a potential true match along a low-dimension manifold. Meanwhile, a neighborhood similarity-based iterative filtering strategy for neighborhood construction is designed to improve the matching performance under the circumstance of seriously deteriorated data. The matching problem is further formulated into a mathematical optimization model based on the neighborhood manifold representation and iterative filtering strategy, and a closed-form solution with linearithmic time complexity (*i.e.*, $O(N\log N)$) is derived, which requires only tens of milliseconds to handle over 1000 putative correspondences. Extensive experiments on general feature matching (F-score $>$ 94% for most cases), remote sensing image registration (RMSE $<$ 3 for most cases), and loop closure detection demonstrate the significant superiority of the proposed method over several state-of-the-art competitors, such as RANSAC, MAGSAC++, and LPM.

1. Introduction

Identifying dependable correspondences between two feature sets is a fundamental problem that usually arises in photogrammetry and computer vision (Ma et al., 2021; Li et al., 2020; Li et al., 2020). This problem is a critical pre-condition for a variety of applications, such as structure-from-motion, SLAM, panoramic stitching, and image and point set registration (Schonberger and Frahm, 2016; Yu et al., 2021; Zhang and Ma, 2021). The matching problem is generally computationally complex due to its combinatorial nature. Specifically, even without taking the outliers into account, establishing one-to-one correspondences between N points and another N points will give rise to $N!$ permutations in total. A popular tactic to tackle the aforementioned challenge is to solve the matching problem in a two-stage mode, *i.e.*, putative set generation and mismatch removal. In the first stage, the tentative set is often generated by simply sorting out local keypoint pairs with sufficiently similar feature descriptors (*e.g.*, scale invariant feature transform, SIFT (Lowe, 2004)). However, in addition to the true correspondences (*i.e.*, inliers), the putative set is typically contaminated by a considerable amount of false correspondences (*i.e.*, outliers). Such contamination is attributed to the ambiguity of local feature representation (especially when the images suffer from inferior-quality, repetitive structures, or occlusion). Thus, devising a robust method is essential in the next step to filter false correspondences for enhancing their dependability.

Many existing methods usually remove false correspondences by imposing a geometric constraint (*e.g.*, motion smoothness constraint), which restricts the correspondences to satisfy an underlying geometrical model. Accordingly, defining a transformation model in advance, which can be either parametric (*e.g.*, affine, homography and epipolar geometry) or non-parametric (*e.g.*, non-rigid), is necessary. Nevertheless, parametric model-based methods readily fail due to an unknown transformation model or the occurrence of non-rigid deformations. Additionally, another drawback of the existing methods is the heavy

* Corresponding author.
E-mail addresses: zizhuo_li@whu.edu.cn (Z. Li), zkn19961212@whu.edu.cn (K. Zhang), shaozhenfeng@whu.edu.cn (Z. Shao), x-gb@163.com (G. Xiao).

https://doi.org/10.1016/j.isprsjprs.2021.11.004
Received 12 August 2021; Received in revised form 3 November 2021; Accepted 3 November 2021
Available online 24 November 2021

computational complexity in non-rigid transformation models. Several techniques recently investigated a local consistent assumption of inliers to conduct mismatch removal (Ma et al., 2019; Bian et al., 2020; Li et al., 2019). The geometric constraint in these approaches is relaxed, followed by filtering outliers via simple yet efficient rules. These approaches can achieve good performance even in scenes undergoing motion discontinuities. However, some gross outliers are preserved as well because underlying intrinsic geometry of the input data is ignored in these techniques.

To address the aforementioned challenges, a robust and efficient approach, which is called *neighborhood manifold representation consensus* (NMRC), is introduced in this paper for robust feature matching. For a pair of images of similar objects or scenes, the spatial neighborhood relationship among feature points characterizing the local topologies is generally stable and may not tend to vary freely owing to physical constraints despite image rotation, scaling, or non-rigid transformation (Ma et al., 2015). Therefore, a mathematical representation of the neighborhoods of feature points is designed via manifold learning, which is similar to local linear embedding (LLE) (Roweis and Saul, 2000), and a mathematical optimization model is introduced. This model removes outliers according to their similarity of the neighborhood manifold representation. The formulation is robust without requiring a pre-defined transformation model. A simple closed-form solution, which has linearithmic time and linear space complexities, is further derived. Moreover, an iterative filtering tactic is devised to provide a relatively clean set of matches for neighborhood manifold representation in the optimization model. Experiments on various real data for general feature matching and two real-world tasks, such as remote sensing image registration and loop closure detection, demonstrate that NMRC can realize more satisfying performance compared with several state-of-the-art alternatives.

The contributions in this paper include the following three aspects:

- Neighborhood manifold representation consensus is presented for feature matching. Compared with a myriad of existing methods that demand a pre-defined global image transformation, the proposed method merely attempts to maintain the consensus of the local topology of the potential inliers using manifold learning. Thus, the proposed method is robust to complex image transformations.
- An iterative filtering strategy is designed for neighborhood construction, which ensures that the neighborhood manifold representation would not be biased by the gross outliers.
- A robust mathematical optimization model is introduced, and its closed-form solution is derived with linearithmic time complexity, which is conducive to many real-time applications. Any metric for characterizing the potential differences between true matches and mismatches can be integrated into the formulation based on this model.

The remainder of this paper is organized as follows. Section 2 describes the background material and related work. Section 3 provides details of the NMRC algorithm for robust feature matching. Section 4 illustrates the experimental results of the proposed approach in comparison with other methods on different types of real-world applications. Section 5 finally presents the conclusion.

2. Related work

Feature matching has been extensively used in various fields, such as computer vision, photogrammetry, remote sensing, and robotics. Several representative reviews regarding this task are summarized in (Ma et al., 2021; Zitova and Flusser, 2003; Jiang et al., 2021). The above-mentioned literature indicates that the existing feature matching methods can be roughly divided into the following three categories: two-step strategy-based, correspondence matrix-based, and learning-based methods.

2.1. Two-step strategy-based methods

In the fields of computer vision, feature descriptors with geometric constraint-based methods typically transform feature matching into a two-step manner (Ma et al., 2014): establishing putative matches and then filtering the false matches using geometrical constraints. The putative set is generally acquired by trimming the set of all possible point correspondences. This scenario is realized by calculating the similarity of feature descriptors at the points and removing the correspondences whose descriptors are overly dissimilar. Representatively, SIFT (Lowe, 2004) compares the distance ratio between the nearest and next-nearest neighbors with a pre-defined threshold to remove unstable matches, demonstrating satisfying performance. Bay et al. (2006) introduced the Haar wavelet calculation to approximate the gradient computation and further accelerate the SIFT operator. In addition, the integral image strategy is applied to simplify computation in the responses of Haar wavelets, enabling its more efficient performance than SIFT. Notably, ORB (Rublee et al., 2011) is currently one of the fastest methods for establishing putative matches. Additionally, the shape context (SC) (Belongie et al., 2002) can be used to construct descriptors when the set of two to-be-matched points has a physical shape. The putative set is inevitably contaminated by a considerable amount of mismatches due to the ambiguity of local descriptors despite various elaborate approaches for putative match generation. Accordingly, the next step is desired to determine and remove the mismatches via geometrical constraints, producing a promising matching performance. To this end, various methods have emerged over the last decades and can be roughly classified into the following four categories: statistical regression, re-sampling, non-parametric fitting, and graph matching methods.

Extensive literature regarding robust estimation is available in the field of statistics (Rousseeuw and Leroy, 2005; Huber, 1981). It has been shown that compared with quadratic L_2 norms, minimizing the L_1 norm is a better choice due to its stronger robustness and capability of tolerating a significant proportion of outliers. The least-median of squares (LMedS) (Rousseeuw and Leroy, 2005) and M-estimators (Huber, 1981) are robust estimators. LMedS can tackle a significant proportion of outliers but is less efficient, while M-estimators require a good initialization of model parameters. Maier et al. (2016) recently presented a statistical optical flow-based guided matching strategy with a favorable performance considering accuracy and efficiency. Re-sampling methods are admittedly the most popular paradigm for this task. These methods follow a hypothesize-and-verify strategy and aim to find the smallest possible outlier-free subset to fit a pre-defined parametric model. Particularly, the random sample consensus (RANSAC) (Fischler and Bolles, 1981) and its variants (*e.g.*, MLESAC (Torr and Zisserman, 2000), GESAC (Li et al., 2020) and MAGSAC++ (Barath et al., 2021)) are the representatives of this type.

Statistical regression and re-sampling methods also have some limitations despite their satisfying results. On the one hand, these methods rely on a geometric parametric model and fail to work under complex transformations. On the other hand, these methods are prone to severe degeneration in the presence of considerable amount of outliers (Li and Hu, 2010). Numerous non-parametric fitting methods have been recently proposed to address the above challenges (Ma et al., 2014; Li and Hu, 2010). These methods interpolate a non-parametric function with the prior condition, in which the motion of the feature correspondence is slow-and-smooth. Nevertheless, the prior condition is not always true in the presence of depth discontinuity or independent motion in the scene. Additionally, these methods typically have high computational complexity, which restricts their applicability to real-time tasks. The graph model provides another view of coping with the feature matching problem. Several representative studies that include spectral matching (Leordeanu and Hebert, 2005), mode-seeking (Wang et al., 2014), and graph shift (GS) (Liu and Yan, 2010) are available. Notably, the transformation model for graph matching is rather flexible. However, this model undergoes similar demerits, namely its non-

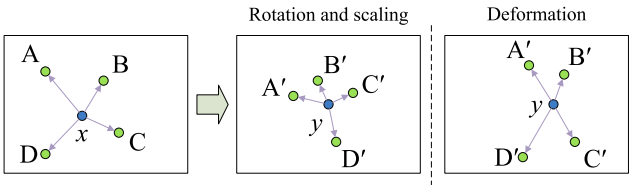


Fig. 1. Illustration of the change of neighborhood topological structure under rotation, scaling and deformation, where (x, y) is a putative correspondence, A, B, C, D are the four nearest neighbors of x and respectively correspond to A′, B′, C′, D′ after transformation.

polynomial-hard nature.

In addition to the aforementioned methods, numerous relaxation methods have been recently investigated. Approaches in this class perform robust estimation of correct matches through the assumption of local structural consistency or piecewise-smoothness constraints. For instance, locality preserving matching (LPM) (Ma et al., 2019) attempts to maintain local neighborhood structures of potential true matches. Local structure consensus constraint (LSCC) (Shao et al., 2020) evaluates each correspondence with a new local structure descriptor to remove mismatches. Grid-based motion statistics (GMS) (Bian et al., 2020) encapsulates the smoothness constraint as a statistic likelihood of a certain number of correspondences in a region and leverages a grid-based scheme to accelerate calculation. Minimum relative motion entropy (MRME) (Shao et al., 2021) translates the feature matching problem into a local relative motion consistency estimation, for which the relative linear motion and the relative angular motion are formulated.

2.2. Correspondence matrix-based methods

Another tactic for feature matching is integrating a correspondence matrix of two feature sets with a parametric or non-parametric geometrical constraint, where the matching problem is also called a pure point set matching problem, *i.e.*, the local-key points typically have no information of feature descriptors. Representatively, iterative closest point (ICP) (Besl and McKay, 1992) utilizes the nearest point strategy to assign a binary correspondence alternatively. Afterward, ICP executes the closed-form rigid transformation estimation through the estimated correspondence until convergence. Chui and Rangarajan (2003) introduced a thin plate spline (TPS)-based general framework for non-rigid matching. Yang et al. (2015) further presented a method, namely global and local mixture distance (GLMD), to enhance the robustness to the data degradation; this method has achieved favorable performance. Many probability-based point set registration methods have emerged in recent years. For example, Myronenko and Song (2010) presented the well-known coherent point drift (CPD) approach based on a Gaussian radial basis function, and many of its variants have been subsequently introduced (Ma et al., 2016; Sun et al., 2020a). These methods cast the matching problem into the estimation of a mixture of densities with Gaussian mixture models (GMM), which are solved within the probabilistic framework and EM algorithm.

Correspondence matrix-based methods can achieve good results in rigid and non-rigid scenarios. However, these methods may fail due to the abundance of outliers or severe data degeneration. In addition, the framework is a complex combinatorial optimization problem with a complex solution space that requires significant time consumption in the iterative estimation process.

2.3. Learning-based methods

In recent years, deep learning methods have been increasingly applied in various fields of computer vision due to their capability of learning and expression and have demonstrated remarkable progress in image matching tasks. A deep learning method is typically applied to learn pixel-level matching relationships directly from image pairs containing the same or similar scene. Concretely, the following three pipelines are popular in current literature: (i) Learning to supersede one or more processes of traditional feature-based methods or directly designing an end-to-end matching network; for example, learning to detect an accurate and reliable set of feature points from an image, learning the main direction or scale of each feature point and its more discriminating and matchable feature descriptor, such as LIFT (Yi et al., 2016) and SuperPoint (DeTone et al., 2018). However, post-processing with a mismatch removal strategy is still required if outliers abound. (ii) Training a convolutional neural network (CNN) to drive an iterative optimization strategy by estimating a similarity metric for two images (Simonovsky et al., 2016). (iii) Directly estimating the transformation parameters via CNN regressors. Representatively, learning to find good correspondences (LFGC) (Moo Yi et al., 2018) was developed as a first attempt for mismatch removal. Given a putative set and the camera intrinsics, LFGC attempts to train a multi-layer perceptron-based deep network to label the correspondences as inliers or outliers and recover the relative pose simultaneously. Nevertheless, LFGC demands known camera intrinsics as input and a specific parameter model, severely limiting its value for practical application. Moreover, LFGC has driven several follow-up works, such as OA-Net (Zhang et al., 2019) and ACNe (Sun et al., 2020b). In addition, SuperGlue (Sarlin et al., 2020) is another novel idea, which is most recently proposed to generate dependable matches from local features with graph neural networks. Nevertheless, this idea still demonstrates a substantial amount of mismatches in the output. Innovatively, Ma et al. (2019) formulated the problem into a two-class classification problem. Such a formulation enables the method to achieve promising matching performance with linearithmic time complexity regarding the data scale but may preserve some gross outliers on account of its match representation defects. Most recently, Chen et al. (2021) introduced a deep learning network called local structure visualization-attention network (LSV-ANet), which aims to transform mismatch removal into a dynamic visual similarity evaluation and achieves favorable performance. However, this idea is sensitive to outliers existing in the small region around a feature point, which limits the ability of LSV-ANet to recognize finegrained patterns and its generalizability for complex matching scenes.

3. Method

The details of the proposed method for robust feature matching are provided in this section. Without loss of generality, a set of N putative matches $S = \{(x_i, y_i)\}_{i=1}^{N}$ is extracted from two given images, where x_i and y_i are 2D column vectors denoting the spatial positions of feature points from two different images, respectively. The following discussion concentrates on removing mismatches from S via neighborhood manifold representation consensus.

3.1. Motivation

Fig. 1 shows that the neighborhood topological structure of the feature point is usually stable and slightly changes under different image transformations. Consequently, exploiting the neighborhood structure information of feature points when dealing with the mismatch removal problem is essential. To this end, a robust neighborhood manifold representation strategy is initially considered for the feature correspondences between two images before presenting the problem formulation.

3.1.1. Neighborhood manifold representation

The neighborhood manifold representation should generally not only capture the neighborhood structure information but also demonstrate robustness to accommodate various kinds of image transformations. To achieve this goal, we design an effective strategy similar to the classical LLE algorithm (Roweis and Saul, 2000), which is

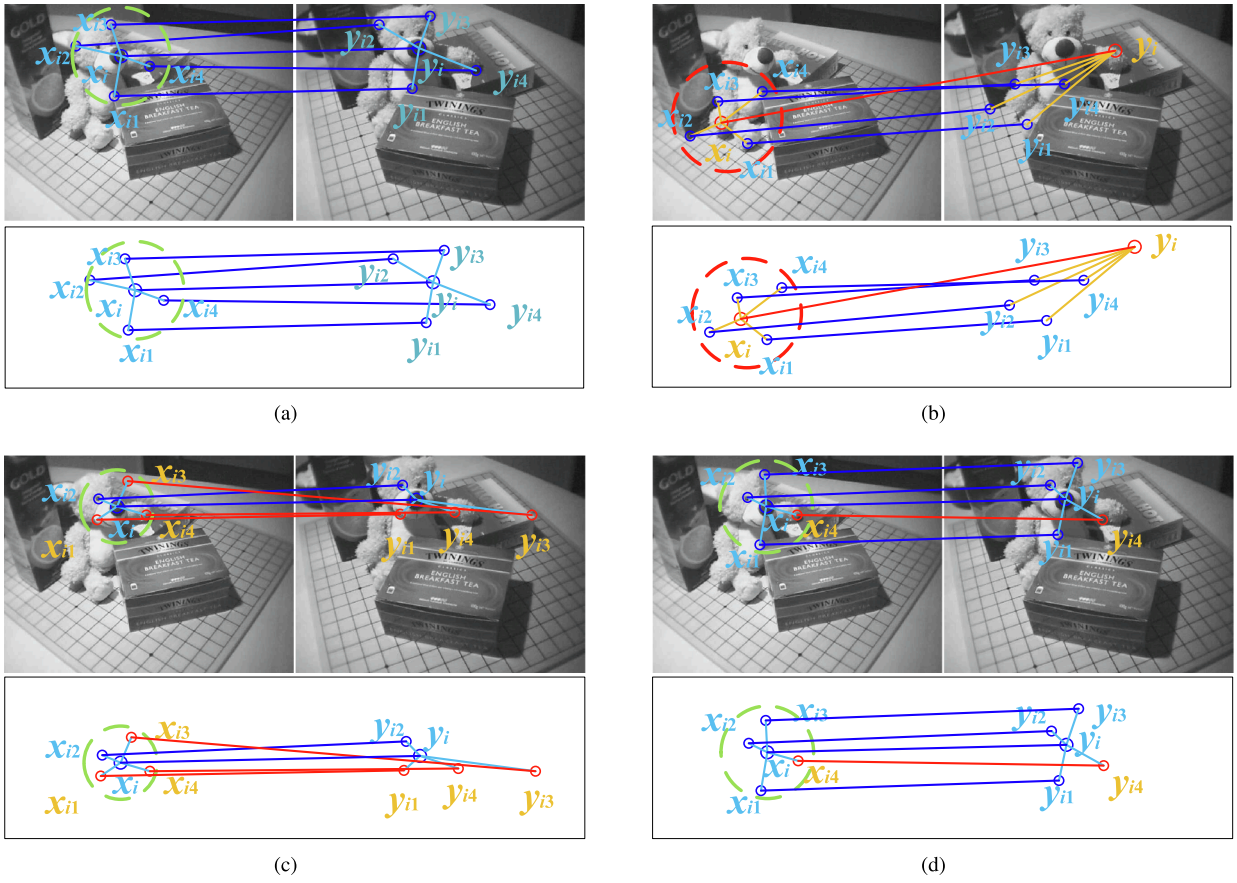


Fig. 2. Illustration of the neighborhood reconstruction consensus measurement. We search the K nearest neighbors ($K = 4$) of feature point x_i and then acquire $\mathcal{N}_{x_i}^K$ which contains $x_{i1}, x_{i2}, x_{i3}, x_{i4}$. Next, $\mathcal{N}_{x_i}^K$ is used to identify $\mathcal{C}_{y_i}^K$ comprised of $y_{i1}, y_{i2}, y_{i3}, y_{i4}$ corresponding to $x_{i1}, x_{i2}, x_{i3}, x_{i4}$, respectively. $\mathcal{N}_{x_i}^K$ and $\mathcal{C}_{y_i}^K$ only contain inliers in the top two cases and they include outliers in the bottom two cases. The distance characterized by Eq. (3) is 0.012, 30.11, 27.55 and 0.109 for the four cases, respectively. Blue: true match; Red: false match. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

presented as a nonlinear dimensionality reduction approach to preserve the local neighborhood structure in a low-dimensional manifold.

In particular, the K-nearest neighbors for each feature point x_i are first searched in $\mathcal{X} = \{x_i\}_{i=1}^N$ and regarded as its neighborhood $\mathcal{N}_{x_i}^K$ under the Euclidean distance. Let $\mathbf{W}^x$ be an $N \times N$ weight matrix and enforce $\mathbf{W}_{ij}^x = 0$ if x_j does not belong to the neighbors of x_i. Second, a reconstruction error measured by the following cost function is minimized:

$$\boldsymbol{\varepsilon}(\mathbf{W}^x) = \sum_{i=1}^{N} \|x_i - \sum_{j=1}^{N} \mathbf{W}_{ij}^x x_j\|^2. \tag{1}$$

Such minimization is performed under a constraint that the rows of the reconstruction weight matrix sum to one: $\forall i, \sum_{i=1}^N \mathbf{W}_{ij}^x = 1$. The optimal weights $\mathbf{W}_{ij}^x$ can be obtained by solving a least-squares problem. Third, $\mathbf{W}^x$ is replaced with an $N \times K$ matrix $\mathcal{W}^x$ by retaining only the non-zero elements in each row of $\mathbf{W}^x$. The robustness to the change in neighborhood topological structure under rotation, scaling, and local deformation can be achieved by using the weight matrix $\mathcal{W}^x$ for neighborhood representation.

Next, this study focused on the construction of the neighborhood representation of each y_i in $\mathcal{Y} = \{y_i\}_{i=1}^N$. To this end, $\{\mathcal{N}_{x_i}^K\}_{i=1}^N$ are used to identify the set $\{\mathcal{C}_{y_i}^K\}_{i=1}^N$, each of which denotes the corresponding feature points of $\mathcal{N}_{x_i}^K$ in $\mathcal{Y}$, as shown in Fig. 2. After acquiring the set $\{\mathcal{C}_{y_i}^K\}_{i=1}^N$, it is combined with Eq. (1), and then each feature point y_i is reconstructed using $\mathcal{C}_{y_i}^K$. Consequently, an $N \times K$ weight matrix $\mathcal{W}^y$ is obtained. The i-th rows of $\mathcal{W}^x$ and $\mathcal{W}^y$ respectively denote the neighborhood topology information of feature points x_i and y_i and can be characterized as:

$$\begin{aligned} \mathcal{W}_i^x &= (\mathcal{W}_{i1}^x, \cdots, \mathcal{W}_{iK}^x) \in \mathbb{R}^{1\times K} \\ \mathcal{W}_i^y &= (\mathcal{W}_{i1}^y, \cdots, \mathcal{W}_{iK}^y) \in \mathbb{R}^{1\times K}. \end{aligned} \tag{2}$$

Concretely, $\mathcal{W}_i^x$ and $\mathcal{W}_i^y$ are $1 \times K$ weight vectors comprising the reconstruction weights of $\mathcal{N}_{x_i}^K$ and $\mathcal{C}_{y_i}^K$, respectively. For instance, $\mathcal{W}_{iK}^x$ denotes the weight of the Kth neighbor in $\mathcal{N}_{x_i}^K$ to reconstruct x_i.

As previously indicated, a correct match should possess a stable local neighborhood structure between two images. Thus, the neighborhood structure information denoted by $\mathcal{W}_i^x$ and $\mathcal{W}_i^y$ should be similar for a correct match (x_i, y_i), as indicated in Fig. 2(a). To this end, a distance metric is defined to characterize the difference in neighborhood topology of (x_i, y_i) as follows:

$$Dist_K(x_i, y_i) = \|\mathcal{W}_i^x - \mathcal{W}_i^y\|^2. \tag{3}$$

Intuitively, the difference in neighborhood topology between the matched feature points (x_i, y_i) is small when the distance in Eq. (3) is close to 0, thus making it an inlier, and vice versa.

For a putative match (x_i, y_i), the above scheme delivers a satisfying indication of the inlier or outlier if $\mathcal{N}_{x_i}^K$ and $\mathcal{C}_{y_i}^K$ are constructed by inliers. As revealed in Figs. 2(a) and (b), $Dist_K(x_i, y_i)$ is 0.012 and 30.11 for an inlier and an outlier with a large margin, respectively. However, the

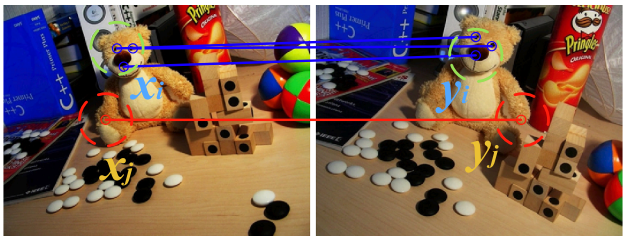


Fig. 3. Schematic illustration of the neighborhood similarity. For a correct match (x_i, y_i), $\mathscr{N}_{x_i}$ should have lots of elements in common with $\mathscr{N}_{y_i}$. For a false match (x_j, y_j), $\mathscr{N}_{x_j}$ ought to have almost no elements in common with $\mathscr{N}_{y_j}$. The blue and red lines represent inliers and outliers, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

inliers cannot be known in advance, and $\mathscr{N}_{x_i}^K$ and $\mathscr{C}_{y_i}^K$ will be inevitably contaminated with outliers when the image undergoes a complex deformation, leading to degenerated performance. As shown in Fig. 2(c), (x_i, y_i) is an inlier, but $Dist_K(x_i, y_i)$ has a large value of 27.55. Therefore, ensuring that $\mathscr{N}_{x_i}^K$ and $\mathscr{C}_{y_i}^K$ contain as few outliers as possible is crucial to boost the matching performance. As revealed in Fig. 2(d), $\mathscr{N}_{x_i}^K$ and $\mathscr{C}_{y_i}^K$ are improved to some extent (still containing an outlier), and $Dist_K(x_i, y_i)$ is then sharply decreased to 0.109. An iterative filtering strategy is introduced below to achieve the aforementioned goal and clean the neighborhoods $\mathscr{N}_{x_i}^K$ and $\mathscr{C}_{y_i}^K$ as much as possible.

3.1.2. Iterative filtering for neighborhood construction

As discussed above, the neighborhood should be sufficiently clean to ensure the distinguishability of $Dist_K$. Therefore, the correspondences used for neighborhood construction should contain a few outliers while retaining the majority of inliers. For an inlier (x_i, y_i) in the putative set S, the neighbors of x_i and y_i should be similar and contain many common elements. On the contrary, the neighbors of x_i and y_i for an outlier (x_i, y_i) should be extremely dissimilar and contain no common elements. This property is known as neighborhood similarity, as shown in Fig. 3. Therefore, given the scale of the neighborhood as κ, the degree of neighborhood similarity between $\mathscr{N}_{x_i}^\kappa$ and $\mathscr{N}_{y_i}^\kappa$ can be characterized as shown below to capture the property mathematically:

$$Ratio(i) = n_i/\kappa, \tag{4}$$

where $n_i = |\mathscr{N}_{x_i}^K \cap \mathscr{N}_{y_i}^K| \leqslant \kappa$ is the number of common elements in the two neighborhoods $\mathscr{N}_{x_i}^\kappa$ and $\mathscr{N}_{y_i}^\kappa$, and $Ratio(i) \in [0, 1]$, which is the ratio of common feature points in corresponding neighborhoods. The value of $Ratio$ induced by an inlier will be large and vice versa. However, $Ratio(i)$ could be small even if (x_i, y_i) is an inlier because S may contain a considerable amount of outliers, and the distinguishability of $Ratio(i)$ will be simultaneously weak when κ is large. An iterative filtering strategy is introduced to address this problem. In particular, a set of thresholds $\{\eta_m\}_{m=1}^M$ is used, and unreliable matches unsatisfying $Ratio > \eta_m$ progressively are filtered. The reliable set $\mathscr{U}_m$ comprises the preserved matches:

$$\mathscr{U}_m = \{(x_i, y_i) \in S | Ratio(i) > \eta_m, i = 1, \cdots, N\}, \tag{5}$$

where $\mathscr{U}_m$ indicates the reliable set after m-th iterations, and η_m indicates the threshold of m-th iterations. Notably, $\{\mathscr{N}_{x_i}^\kappa\}_{i=1}^N$ and $\{\mathscr{N}_{y_i}^\kappa\}_{i=1}^N$ are constructed using $\mathscr{U}_{m-1}$ when $\mathscr{U}_m$ is obtained by Eqs. (4) and (5), and $\mathscr{U}_0$ is defined as S. The reliable set is also sought on the basis of the entire set S rather than $\mathscr{U}_{m-1}$. In addition, the initial threshold η_1 is set to be a small value. Therefore, $\mathscr{U}_1$ contains sufficient correspondences for neighborhood construction in the next iteration. The final reliable set $\mathscr{U}_M$ is typically sufficiently clean for neighborhood construction during distance metric calculation in Eq. (3) by using this iterative filtering strategy.

3.2. Problem formulation

The mismatch removal problem is formulated into a mathematical optimization model in this section based on the proposed robust neighborhood manifold representation. In particular, the optimal solution to preserve local topology of feature points is

$$\mathscr{I}^* = \arg\min_{\mathscr{I}} C(\mathscr{I}; S, \mathscr{U}_M, \lambda), \tag{6}$$

with C denoting the cost function defined as:

$$C(\mathscr{I}; S, \mathscr{U}_M, \lambda) = \sum_{i \in \mathscr{I}} Dist_K(x_i, y_i) + \lambda(N - |\mathscr{I}|), \tag{7}$$

where $Dist_K(x_i, y_i)$ defined in Eq. (3) measures the consensus of neighborhood manifold representation between x_i and y_i, the reliable set $\mathscr{U}_M$ is used to construct $\{\mathscr{N}_{x_i}^K\}_{i=1}^N$ and $\{\mathscr{C}_{y_i}^K\}_{i=1}^N$, and $|\cdot|$ is the cardinality of a set. The first term in this cost function penalizes any correspondence not preserving the consensus of neighborhood manifold representation, the second term is utilized to discourage the mismatches, and parameter $\lambda > 0$ is used to balance the two terms. The optimal solution should ideally realize zero penalty, *i.e.*, the first term of C should be zero. This solution attempts to minimize the value of the cost function while acquiring the largest set of inliers. Notably, the objective function in Eq. (6) is translation, rotation, and scaling invariant. The issue of removing mismatches can then be solved by minimizing Eq. (6).

An $N \times 1$ binary vector $\mathbf{p}$ is introduced to optimize the objective function (6) and characterize the correctness of the putative correspondences, *i.e.*, $p_i = 1$ for an inlier and 0 for an outlier. Accordingly, Eq. (7) can be written as follows:

$$C(\mathbf{p}; S, \mathscr{U}_M, \lambda) = \sum_{i=1}^N p_i Dist_K(x_i, y_i) + \lambda(N - \sum_{i=1}^N p_i). \tag{8}$$

Its form is then reorganized by merging the terms related to p_i and obtain:

$$C(\mathbf{p}; S, \mathscr{U}_M, \lambda) = \sum_{i=1}^N p_i(c_i - \lambda) + \lambda N, \tag{9}$$

where

$$c_i = Dist_K(x_i, y_i) \tag{10}$$

measures whether the i-th correspondence (x_i, y_i) fulfills the geometric constraint of preserving the neighborhood reconstruction consensus. An inlier will yield zero or a small cost, while an outlier will lead to a large cost.

3.3. A closed-form solution

The local topology between the feature points is fixed given a tentative set S. Hence, $\{c_i\}_{i=1}^N$ can be estimated beforehand. The only unknown variable in Eq. (9) is p_i. The estimation of the value of p_i indicates that any putative match with a cost smaller than λ will result in a negative term and decrease the objective function. Thus, setting the value of p_i to 1 is preferred, and vice versa. Specifically, the optimal solution of $\mathbf{p}$ that minimizes Eq. (9) can be determined by the following simple criterion:

$$p_i = \begin{cases} 1, & \text{if } c_i < \lambda \\ 0, & \text{if } c_i > \lambda \end{cases}, \quad i = 1, 2, \ldots, N. \tag{11}$$

Consequently, the optimal inlier set $\mathscr{I}^*$ can be represented as:

$$\mathscr{I}^* = \{i | p_i = 1, i = 1, 2, \ldots, N\}. \tag{12}$$

The parameter λ acts as a threshold to determine the correctness of the putative set. Simultaneously, the neighborhood $\mathscr{N}_x$ is constructed on the basis of the reliable set $\mathscr{U}_M$ due to the iterative filtering strategy. Thus, the value of $Dist_K(x_i, y_i)$ is an accurate indication of the consensus of local topology between two corresponding feature points x_i and y_i.

The above formulation can generate satisfactory results. However, constructing the local neighborhoods $\{\mathscr{N}_{x_i}^K\}_{i=1}^N$ based on an ideal inlier set $\mathscr{I}$ will further promote the outlier removal performance, especially when the putative set S abounds with outliers. The matching accuracy of the match set obtained by Eq. (12) is considerably higher than that of $\mathscr{U}_M$. Therefore, after obtaining an inlier set $\mathscr{I}_0$ based on Eq. (12), *i.e.*, $\mathscr{I}_0 = \text{argmin}_{\mathscr{I}} C(\mathscr{I}; S, \mathscr{U}_M, \lambda)$, we use it to replace $\mathscr{U}_M$ for neighborhood construction. Finally, the optimal inlier set $\mathscr{I}^*$ is obtained as follows:

$$\mathscr{I}^* = \underset{\mathscr{I}}{\text{argmin}} C(\mathscr{I}; S, \mathscr{I}_0, \lambda). \tag{13}$$

Notably, the above process can be iterated to further promote the matching performance, *i.e.*, iteratively using the correspondence set generated in the previous iteration for neighborhood construction until convergence. However, such an iteration can only slightly improve the matching performance, which indicates that $\mathscr{I}_0$ is sufficiently good to approximate the true inlier set for neighborhood construction. Thus, only Eq. (13) is used to determine the optimal inlier set $\mathscr{I}^*$. The proposed approach focuses on preserving the consensus of neighborhood manifold representation. Therefore, this approach is named NMRC, and the entire procedure is summarized in Alg. 1.

Algorithm 1. The NMRC Algorithm

Input: Putative set $S = \{(x_i, y_i)\}_{i=1}^N$, parameters $K, \kappa, \{\eta_m\}_{m=1}^M, \lambda$
Output: Inlier set $\mathscr{I}^*$
1 Initialize $m = 0, \mathscr{U}_0 = S$
2 **Iteration**:
3 $m = m + 1$
4 Construct neighborhoods $\{\mathscr{N}_{x_i}^\kappa, \mathscr{C}_{y_i}^\kappa\}_{i=1}^N$ using $\mathscr{U}_{m-1}$
5 Determine $\mathscr{U}_m$ using Eqs. (4) and (5)
6 **Until**: $m \geqslant M$ 7 Construct neighborhoods $\{\mathscr{N}_{x_i}^K, \mathscr{C}_{y_i}^K\}_{i=1}^N$ using $\mathscr{U}_M$
8 Calculate $\{c_i\}_{i=1}^N$ using Eqs. (3) and (10)
9 Determine $\mathscr{I}_0$ using Eqs. (6), (11) and (12)
10 Construct neighbors$\{\mathscr{N}_{x_i}^K, \mathscr{C}_{y_i}^K\}_{i=1}^N$ using $\mathscr{I}_0$
11 Calculate $\{c_i\}_{i=1}^N$ using Eqs. (3) and (10)
12 Determine $\mathscr{I}^*$ using Eqs. (13), (11) and (12).

3.4. Discussion

Notably, albeit the proposed method has a similar formulation to the work of LPM (Ma et al., 2019), there are significant differences between them. First, LPM constructs the neighborhood of a feature point x just using its K nearest neighbors in putative set S, *i.e.*, the putative set S is used to construct $\{\mathscr{N}_{x_i}^K\}_{i=1}^N$. When S contains a great deal of outliers, the neighborhood will be inevitably contaminated with outliers, resulting in degenerated performance of LPM. By contrast, the proposed method utilizes a neighborhood similarity-based iterative filtering strategy for neighborhood construction, *i.e.*, the reliable set $\mathscr{U}_M$ is used to construct $\{\mathscr{N}_{x_i}^K\}_{i=1}^N$, which can boost the matching performance. Second, the neighborhood structure similarity between two feature points in LPM only considers the simple motion statistics of K-NN and cannot exploit the true local structure. In other words, the measuring criterion in LPM is not strict enough to preserve the neighborhood topological structure. However, the proposed method introduces a more rigorous constraint for neighborhood topological structure preservation by leveraging manifold learning to exploit fully the underlying intrinsic geometry information of feature correspondences. In a nutshell, the proposed method is capable of preserving neighborhood topological consensus more strict with more robust and accurate matching performance. This will be validated in the subsequent experimental results.

3.5. Computational complexity

Given the putative set S comprising N correspondences, the time complexity of searching K nearest neighbors for each correspondence in S is close to $O((K+N)\log N)$ by using K-D tree (Bentley, 1975). Thus, the time complexity of Lines 2–6 and Line 8 in Algorithm 1 is about $O(M(\kappa+N)\log N)$ and $O((K+N)\log N)$, respectively. Calculating the cost $\{c_i\}_{i=1}^N$ in Lines 9 and 12 requires obtaining the reconstruction weight matrix $\mathbf{W}$, which has $O(K^3N)$ time complexity according to Eq. (1) because each row of $\mathbf{W}$ can be solved separately with $O(K^3)$ time complexity. Considering the similarity of time complexity in Lines 11 and 8, the total time complexity of the proposed NMRC is about $O((M\kappa + MN + K + N)\log N + K^3N)$. The space complexity of NMRC is $O((\kappa+K)N)$ considering the memory for storing the neighborhoods $\{\mathscr{N}_{x_i}^\kappa\}_{i=1}^N, \{\mathscr{N}_{y_i}^\kappa\}_{i=1}^N, \{\mathscr{N}_{x_i}^K\}_{i=1}^N$, and $\{\mathscr{C}_{y_i}^K\}_{i=1}^N$. M, K, and κ are generally far smaller than N. Hence, the time and space complexities of the approach can be simply written as $O(N\log N)$ and $O(N)$, respectively. Therefore, NMRC has linearithmic time and linear space complexities in regard to the scale of the given tentative set. Consequently, the proposed method is suitable for tackling real-world tasks.

3.6. Implementation details

Several parameters have to be set for NMRC: $K, \kappa, \{\eta_m\}_{m=1}^M, M$, and λ. Parameter K represents the number of nearest neighbors for neighborhood representation based on manifold learning. Parameter κ controls the size of the neighborhood construction in the iterative filtering strategy. Parameter η_m is a threshold, which is used for determining whether a correspondence belongs to the reliable set. Parameter M determines the iteration times of the iterative filtering strategy. Parameter λ acts as the threshold for distinguishing the correctness of putative matches. Large values of M, η_m, and κ will generally improve the quality of reliable correspondence set but sacrifice a portion of true matches simultaneously. A large value of K and a small value of λ will increase the precision and simultaneously decrease the recall, and vice versa. The default values in the experimental evaluation are set as $K = 10, \kappa = 10$, $\eta_m = [0.2, 0.5, 0.5], M = 3$, and $\lambda = [0.12, 0.12]$, and the practicality of the parameter settings is verified through a series of ablation studies in the next section.

4. Experimental results

The ablation experiments of the proposed NMRC are performed in this section, and the performance on feature matching for real image pairs is then evaluated. Subsequently, the robustness of NMRC is tested, and this method is finally applied to tackle two real-world tasks, *i.e.*, image registration and loop closure detection. The open source VLFeat is used to search the K-nearest neighbors with K-D tree (Vedaldi and Fulkerson, 2010). All the experiments are conducted on a desktop with 2.4 GHz Intel Core i5-6200U CPU, 8 GB memory, and MATLAB code.

4.1. Ablation studies

A total of 30 real image pairs with different kinds of transformations containing rigid, rotation, scale change, and non-rigid deformation are selected for evaluation to validate the effectiveness of the proposed method. The average initial inlier percentage of putative correspondences obtained by SIFT on the entire test data is approximately 57.02%, which makes the mismatch removal task challenging. The precision and

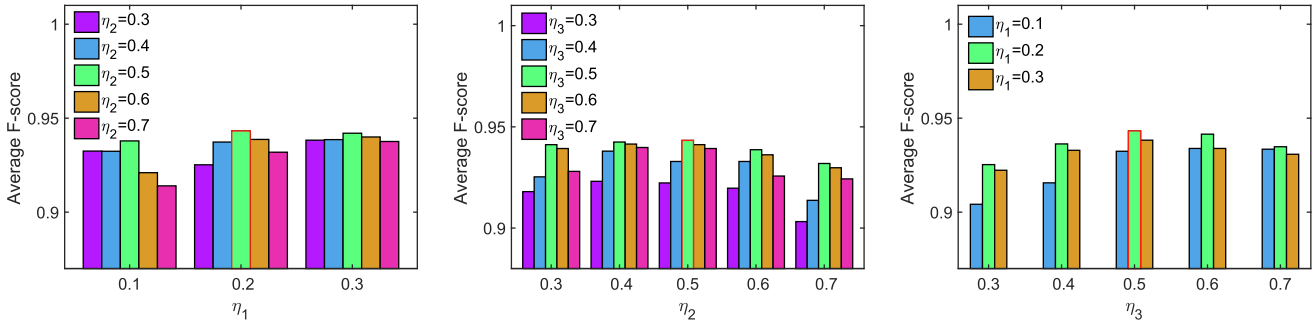


Fig. 4. Average F-score of our method with different parameter settings of $\{\eta_m\}_{m=1}^{M}$ on the 30 test image pairs. From left to right, we respectively fix η_3, η_1 and η_2 and change the other two to find the optimal threshold set $\{\eta_m\}_{m=1}^{M}$.

Table 1
Average F-score (AF) and average runtime (ART, unit: ms) of our method in regard to different parameter κ on the 30 test image pairs in Fig. 4. Bold indicates the best.

κ	6	8	10	12	14
AF	94.04%	94.21%	**94.33%**	93.01%	91.98%
ART	44.9	45.1	45.5	46.2	46.6

Table 2
Average F-score (AF) and average runtime (ART, unit: ms) of our method in regard to different parameter K on the 30 test image pairs in Fig. 4. Bold indicates the best.

K	6	8	10	12	14
AF	93.59%	93.95%	**94.33%**	93.39%	92.57%
ART	40.3	42.3	45.5	48.1	51.1

Table 3
Average F-score (AF) and average runtime (ART, unit: ms) of our method in regard to different parameter M on the 30 test image pairs in Fig. 4. Bold indicates the best.

M	1	2	3	4
AF	92.15%	92.28%	**94.33%**	94.27%
ART	43.4	44.5	45.5	47.8

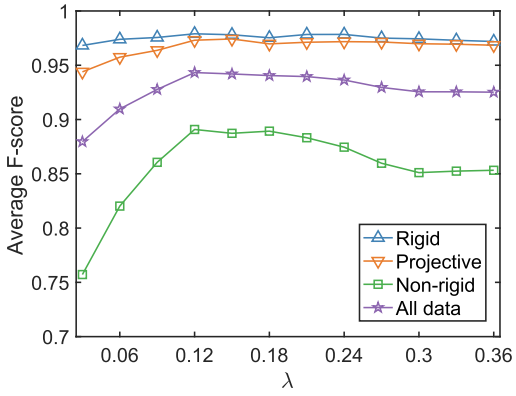


Fig. 5. Average F-score of our method with different settings of λ on various types of image transformations that are classfied from the selected test image pairs.

recall are adopted to characterize the performance; precision is defined as the ratio of the identified inlier number and the preserved match number, and recall is the ratio of identified inlier number and the entire inlier number.

First, the best parameter settings of η_m, κ, K, and λ are investigated. Consequently, the average F-scores are tested with regard to different parameter settings of η_m on the 30 test image pairs, and the results are presented in Fig. 4. In particular, one parameter of $\{\eta_1, \eta_2, \eta_3\}$ is fixed as its "optimal" parameter setting, and the two other parameters are changed to determine the optimal threshold set $\{\eta_m\}_{m=1}^{M}$. The results indicate that $\eta_m = [0.2, 0.5, 0.5]$ realizes the best average F-score, which is regarded as the default optimal η_m throughout this paper. Different settings are also tested on the 30 test image pairs to identify the optimal values of κ, K, and M, and the average F-scores and runtime are summarized in Tables 1–3. The results indicate that $\kappa = 10, K = 10$, and $M = 3$, which are considered to be their default optimal settings, can achieve the best performance. Considering the strong influence of parameter λ, the average F-scores are tested using different λ on various types of image transformations, which are classified from the selected test image pairs, and the results are reported in Fig. 5. These results reveal that the best choice of λ may slightly change considering various types of image transformations. Therefore, $\lambda = 0.12$ is set as its default optimal setting due to its best average F-score on all data. Note that the optimal parameters can be determined more accurately by using a grid search approach. However, the large number of parameters in the proposed method (*e.g.*, more than 6) makes it difficult to use the grid search.

Then, the effectiveness of the iterative filtering strategy for neighborhood construction is substantiated. Precision and recall curves on the 30 image pairs are summarized in Fig. 6, wherein the results of NMRC without and with the iterative filtering strategy in the case of different values of λ are reported. The iterative filtering strategy significantly enhances the matching performance. The proposed method preserves approximately 93.33% of the true matches with an appropriate value of λ (*e.g.*, 0.12), and the precision can reach up to 95.74% simultaneously.

The advantages of utilizing $\mathscr{I}_0$ for neighborhood construction are also considered. To this end, the distributions of the cost c_i are reported in Eq. (10) by utilizing S, $\mathscr{U}_M$ and $\mathscr{I}_0$ for neighborhood construction in Fig. 7. Several outliers would generally be observed in an inlier neighborhood when the entire putative set is used to construct the neighborhoods due to the existence of contaminated data, as shown in Fig. 2 (c). Thus, no clear dividing line is found between the distribution of inliers and outliers. Fortunately, using the reliable $\mathscr{U}_M$ to construct neighborhoods can enlarge the margin between inliers and outliers significantly. Nevertheless, the margin between inliers and outliers has been further enlarged by using the refined $\mathscr{I}_0$ for neighborhood construction. The average precision and recall on the 30 test image pairs can be further increased from $(95.74\%, 93.33\%)$ to $(96.84\%, 97.03\%)$ with the parameter $\lambda = 0.12$. Notably, iteratively using Eq. (13) for neighborhood construction until convergence can only slightly increase the average precision-recall pair to $(97.12\%, 97.34\%)$. However, this approach is time consuming.

Finally, as Eq. (13) also plays a role of iterative filtering, the use of $\mathscr{I}_0$ without $\mathscr{U}_M$ is considered for neighborhood construction. That is to

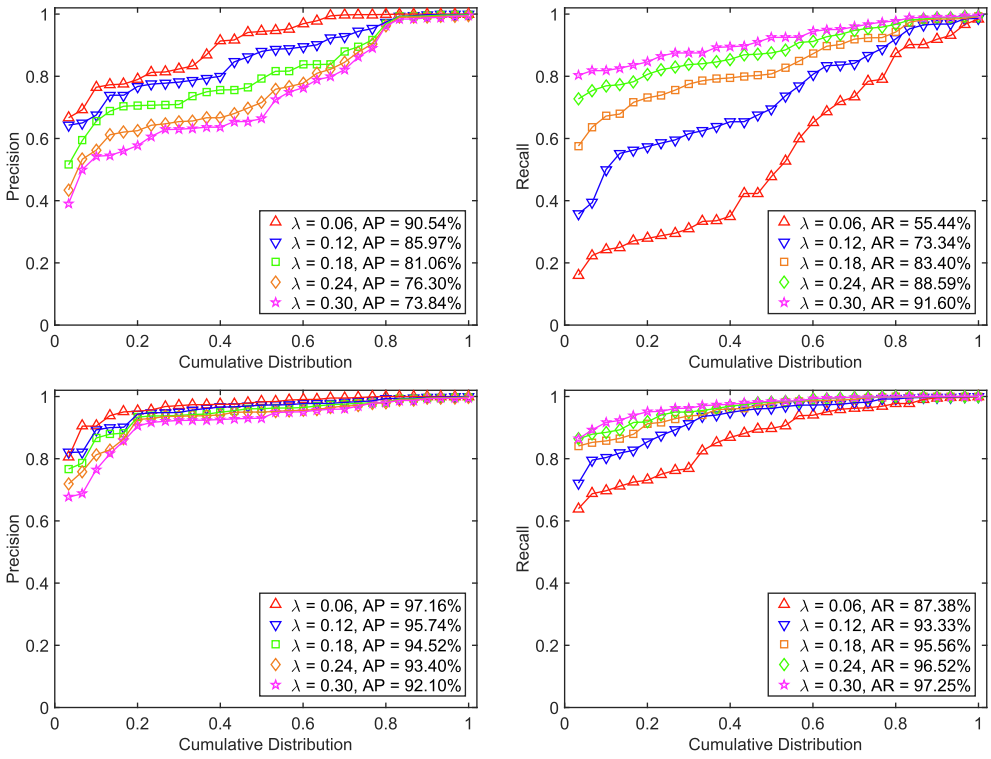


Fig. 6. Precision and recall with respect to the cumulative distribution by using the whole set S (top) and the reliable set $\mathscr{U}_M$ (bottom) to construct neighborhood $\{\mathscr{N}_{x_i}^K\}_{i=1}^N$ and $\{\mathscr{N}_{y_i}^K\}_{i=1}^N$ on 30 image pairs using different λ. A point on the curve with coordinate (x, y) denotes that there are $100 * x$ percents of image pairs which have precision or recall no more than y. AP: average precision; AR: average recall.

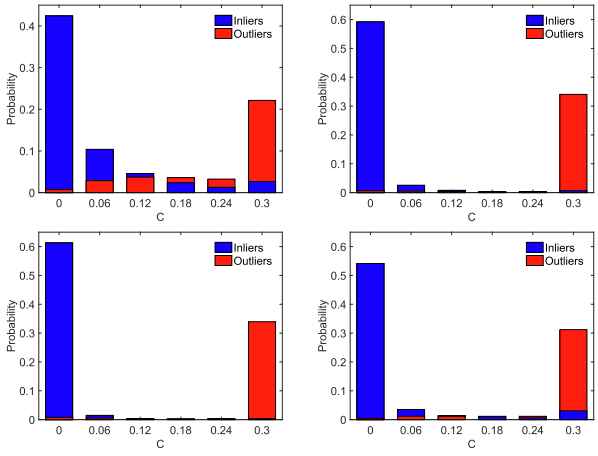


Fig. 7. Distribution of cost c_i in Eq. (10) by using the whole set S (top left), $\mathscr{U}_M$ (top right), $\mathscr{I}_0$ (bottom left), and $\mathscr{I}_0$ without $\mathscr{U}_M$ (bottom right) to construct neighborhood structure $\{\mathscr{N}_{x_i}^K\}_{i=1}^N$ and $\{\mathscr{C}_{y_i}^K\}_{i=1}^N$.

say, S is utilized rather than $\mathscr{U}_M$ for neighborhood construction to obtain $\mathscr{I}_0$, and then $\mathscr{I}^*$ in Eq. (13) is estimated. The result is reported in the bottom right plot of Fig. 7, where the precision-recall pair is only $(95.31\%, 78.80\%)$. Even after iteratively using Eq. (13) for neighborhood construction until convergence, the final precision-recall pair is $(93.28\%, 83.57\%)$, which is substantially less than the result $(96.84\%, 97.03\%)$ when using $\mathscr{I}_0$ with $\mathscr{U}_M$ for neighborhood construction. Therefore, the proposed iterative filtering strategy is essential for promoting the matching performance in the model.

4.2. Results on feature matching

4.2.1. Qualitative illustration

Ten representative image pairs are selected to provide an intuitive illustration of the matching performance of the NMRC, and the results are presented in Fig. 8. These image pairs involve different kinds of transformations, including affine (1st), nonrigid deformation (2nd, 3rd, 4th, and 5th), and epipolar geometry (6th, 7th, 8th, 9th, and 10th). For the results of each group, the left plot schematically demonstrates the matching result, and the right plot is the vector field representation of putative matches. Precision, recall, and F-score are used to characterize the matching performance, where the F-score is defined as the ratio of $2 \times \text{Precision} \times \text{Recall}$ and $\text{Precision} + \text{Recall}$. Each tentative match in each image pair is manually examined to establish the ground truth, and the benchmark is provided before conducting experiments to ensure its objectivity. With the NMRC, the statistics of precision, recall, and F-score on the 10 image pairs are as follows: (99.66%, 100.0%, 99.83%), (99.57%, 96.23%, 97.87%), (97.68%, 96.99%, 97.34%),(99.01%, 100.0%, 99.50%), (94.83%, 90.66%, 92.30%),(96.67%, 90.14%, 93.43%), (98.35%, 100.0%, 99.17%),(98.45%, 99.22%, 98.83%), (95.71%, 95.71%, 95.71%), and (100.0%, 100.0%, 100.0%). These results reveal that most true correspondences are successfully determined with only very few misjudgments on all test image pairs. This finding also validates that the NMRC is sufficiently robust to tackle various kinds of transformations despite the abundance of outliers.

4.2.2. Quantitative comparison

Experiments are conducted on five representative datasets (Ma et al., 2021), namely *Daisy*, *DTU*, *Adelaide* (Wong et al., 2011), *RS*, and *Retina*, to provide a comprehensive performance comparison. Specifically, *Daisy* comprises wide baseline image pairs with ground-truth depth maps, including two short image sequences and several individual image pairs, from which a total of 52 image pairs are created for evaluation. *DTU*

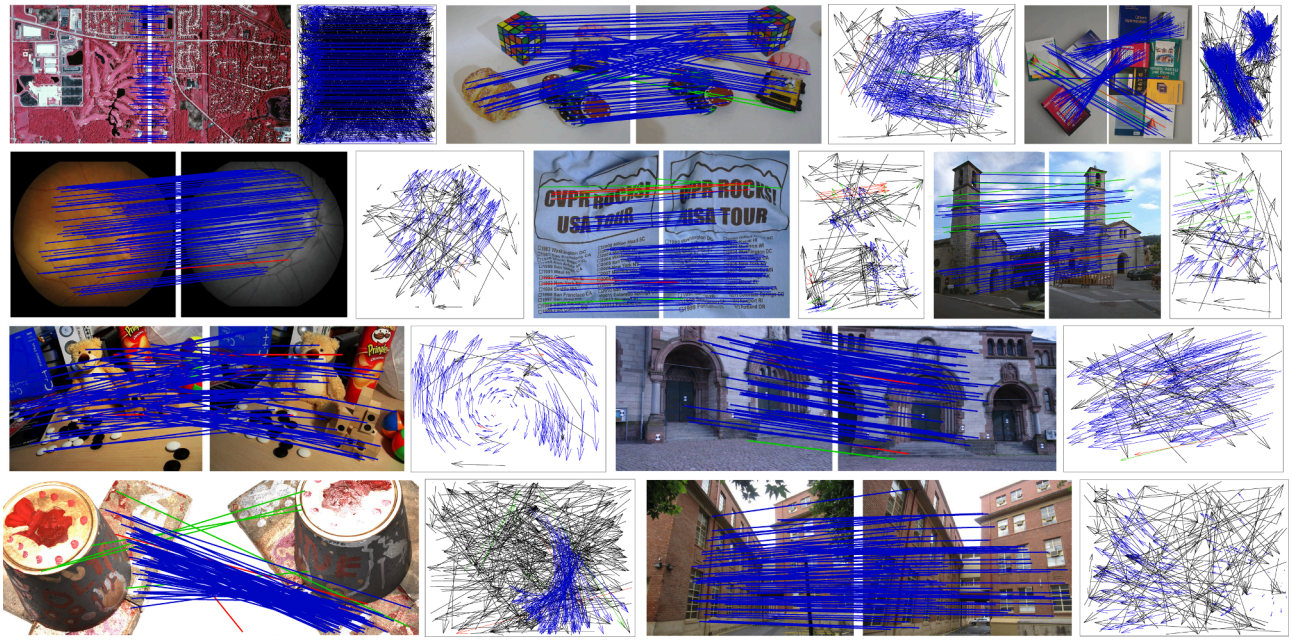


Fig. 8. Feature matching results of our NMRC on 10 representative image pairs. From top to bottom and left to right: *Land*, *Cubebreadtoychips*, *Book*, *Retina*, *T-shirt*, *Church*, *Bear*, *Herzjesu*, *Frustum* and *Sene*. The number of putative matches on the 10 image pairs are 2203, 327, 746, 183, 300, 126, 198, 196, 400 and 250, respectively. The ratio of inliers in the 10 image pairs are 13.21%, 73.09%, 75.74%, 54.64%, 60.67%, 56.35%, 90.40%, 65.31%, 31.25% and 52.80%, respectively. 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, red = false positive). For visibility, in the image pairs, at most 100 randomly selected matches are presented, and the true negatives are not shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

includes a large number of different scenes with ground-truth camera positions, from which two scenes (*i.e.*, *Frustum* and *House*) are chosen to create 131 image pairs involving large viewpoint changes for evaluation. *Adelaide* encompasses 38 image pairs, including a number of image pairs of different buildings, which is mainly related to epipolar geometry, and several image pairs with multiple motions, *i.e.*, multiple objects exist and undergo independent movements between two images. *RS* is a remote sensing dataset comprising 161 image pairs, including color-infrared, SAR, and panchromatic photographs. *Retina* is a medical image dataset containing 70 retinal image pairs involving non-rigid transformations.

The correctness of feature correspondence for the first three publicly available datasets in a putative set is identified on the basis of ground truth provided by the corresponding datasets. The ground-truth correspondences for the remaining datasets are established considering a benchmark as previously mentioned. Particularly, the correctness of each feature correspondence of each image pair is manually examined.

Nine state-of-the-art methods for robust feature matching are adopted for comparison: RANSAC (Fischler and Bolles, 1981), MLESAC (Torr and Zisserman, 2000), MAGSAC++ (Barath et al., 2021), ICF (Li and Hu, 2010), GS (Liu and Yan, 2010), GMS (Bian et al., 2020), LPM (Ma et al., 2019), mTopKRP (Jiang et al., 2019), and LFGC (Moo Yi et al., 2018). These alternatives cover different categories in the literature, and thus are proper delegates of the field of feature matching. It should be mentioned that RANSAC, MLESAC and MAGSAC++ use a homography model and the iteration number is set to 10000 for balancing performance and efficiency. For RANSAC and MLESAC, the threshold for determining an inlier is set to 3 pixels. As for MAGSAC++, σ_{max} is set considering 50 pixels, which is suggested by the original paper. The remaining methods are implemented on the basis of original papers, and the researchers of this study attempted to tune their parameters to realize their best performance. Moreover, the parameters of the nine methods are fixed throughout the experiments.

The initial inlier ratio, precision, recall, F-score, and runtime statistics on the five datasets are summarized in Fig. 9. Notably, the inlier ratio in the fourth dataset is quite low, which complicates the mismatch removal. The average number of tentative correspondences on the five datasets is approximately 1475.60, 544.99, 217.18, 445.34, and 69.03. For the matching performance, RANSAC has a satisfying performance considering precision and recall on *Daisy*, *DTU*, *RS*, and *Retina*. This performance may be due to the sufficient sampling times spent to acquire an outlier-free subset to estimate transformation despite the low inlier ratio. MLESAC can consistently achieve the best matching precision on all datasets but performs poorly considering recall on *Daisy*, *DTU*, and *Adelaide*. Such a performance is caused by leveraging the maximum likelihood process to verify the model quality instead of only the number of inliers, which can improve precision but sacrifice recall simultaneously. MAGSAC++ performs favorably considering recall on *RS* and *Retina* but achieves poor performance considering precision on *Retina*. Notably, RANSAC, MLESAC and MAGSAC++ all perform poorly on *Adelaide* because they cannot handle severe non-rigid transformations. ICF usually has a promising performance considering precision or recall, but not at the same time. This result is probably due to the assumption of ICF that the motion field is globally smooth, which fails to tackle the scenes with large depth discontinuity or motion inconsistency. GS achieves satisfying performance on *Adelaide* but performs poorly on the remaining datasets because it cannot estimate the factor of affinity matrix automatically and its non-affine invariance. The performance of GMS is unsatisfactory, especially on the *Retina* dataset, because it was originally designed with a considerable amount of tentative correspondences to realize improved performance. LFGC typically yields results of satisfactory precision. However, the recall is quite low because it mainly aims to identify good correspondences and accurately recover the relative pose simultaneously. Therefore, LGFC may falsely remove a set of unstable true correspondences, yielding to a low recall. Furthermore, several testing data involving non-rigid deformations or low-overlapped areas are rather different from the training data of LFGC, which usually undergo large-scale or viewpoint changes. Additionally, LFGC demands camera intrinsics as input to normalize data, which are unavailable in the presented testing data.

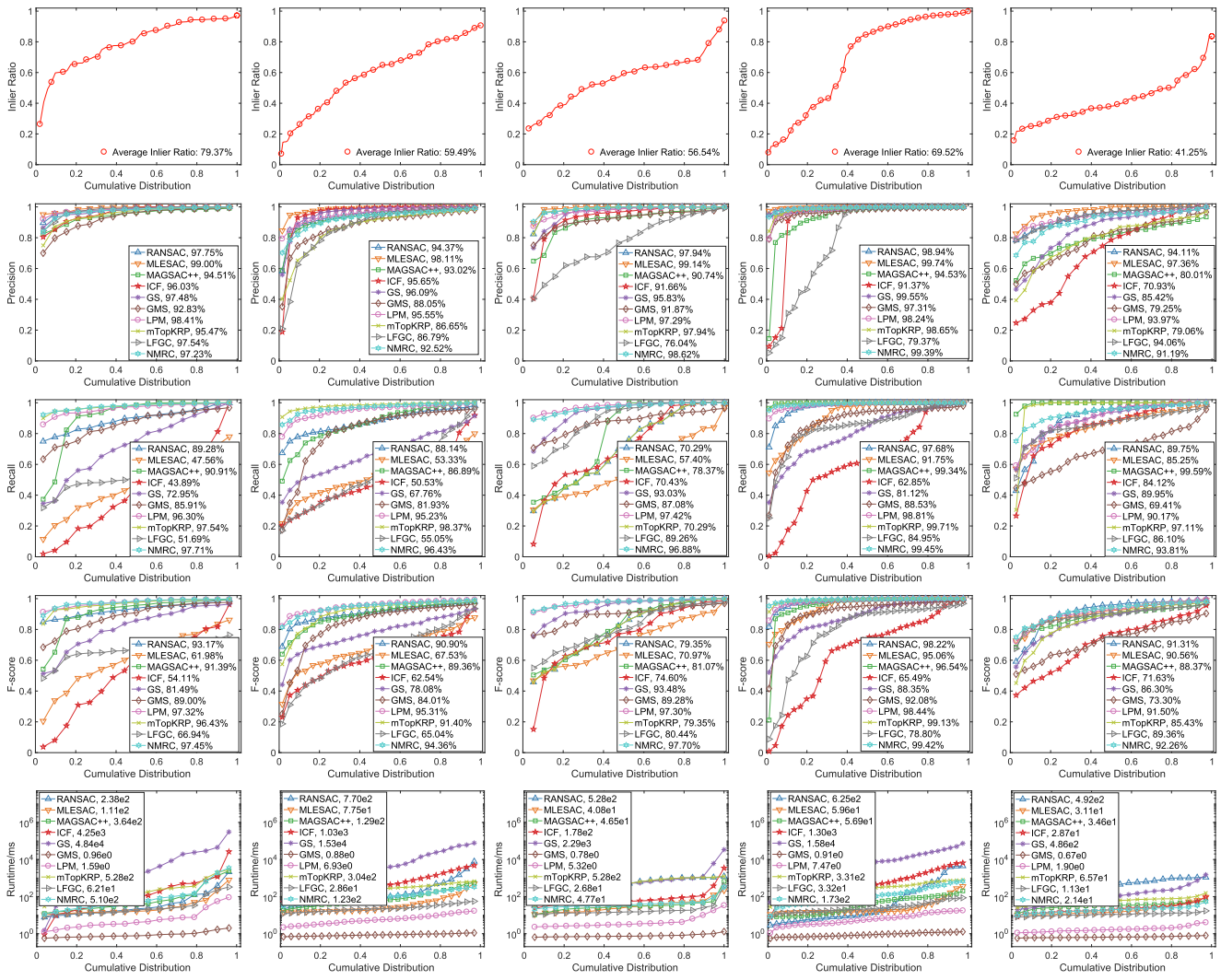


Fig. 9. Quantitative comparisons of RANSAC, MLESAC, MAGSAC++, ICF, GS, GMS, LPM, mTopKRP, LFGC and NMRC on five datasets, such as (left to right) *DAISY*, *DTU*, *Adelaide*, *RS*, and *Retina*. (Top to Bottom) Initial inlier ratio, precision, recall, F-score and run time with respect to the cumulative distribution.

LPM typically has a favorable performance considering precision and recall simultaneously due to its capability of addressing various kinds of transformations. mTopKRP has a rather promising performance on *Daisy* and *RS* but performs poorly on the remaining datasets. Such a performance is due to the effective preservation of the neighborhood topology by the ranking list distance measurement when the transformation between the image pairs is relatively simple. However, mTopKRP will be no longer workable in the case of complex non-rigid transformations. Compared with the aforementioned methods, NMRC can trade-off precision against recall (*i.e.*, F-score) favorably, because this approach is dispensed with a motion model between image pairs and fully captures the underlying intrinsic geometry of the input data. Additionally, the time costs of these methods are shown in Fig. 9, indicating that NMRC can efficiently eliminate mismatches.

4.2.3. Robustness test

This section further verifies the performance of NMRC under different deformation degrees and provides a comparison with the eight state-of-the-art methods. In particular, the proposed method is tested on a group of images with five different deformation degrees, which contains 8 scenes selected from *Daisy*, *DTU*, and *VGG* (Ma et al., 2021) in Fig. 10. Notably, the images from left to right are deemed as an increasing degree of deformation compared with those of the first column. The first image in each row is paired with the remaining images in sequence to generate five test image pairs for each scene. The average value and standard deviation of inlier rate and F-score of the aforementioned methods regarding the deformation degree are shown in Fig. 11. These results indicate that the performances of all methods degenerate with the increase in the degree of deformation, but the proposed method changes marginally. Thus, NMRC possesses strong robustness and outperforms the other competitors.

4.3. Results on remote sensing image registration

Image registration is one of the most important applications for image feature matching. It focuses on whether the warped sensed image enables maximization of the alignment of the overlapping area between the reference and sensed images. Therefore, the dependable feature matches are first acquired using NMRC. Afterward, thin plate spline (TPS) is chosen to generate a smooth fitting for the feature matches by estimating the transformation function $\mathcal{F}$ considering its generality and smooth functional mapping in supervised learning (Chui and Rangarajan, 2003). Consequently, such an approach can address the non-rigid transformation in feature matching. Furthermore, TPS has no free parameters that require manual tuning, and its closed-form solution can be decomposed into a global linear affine motion and a local non-affine warping component. Finally, each pixel in the sensed image is mapped to the corresponding coordinate with the estimated transformation

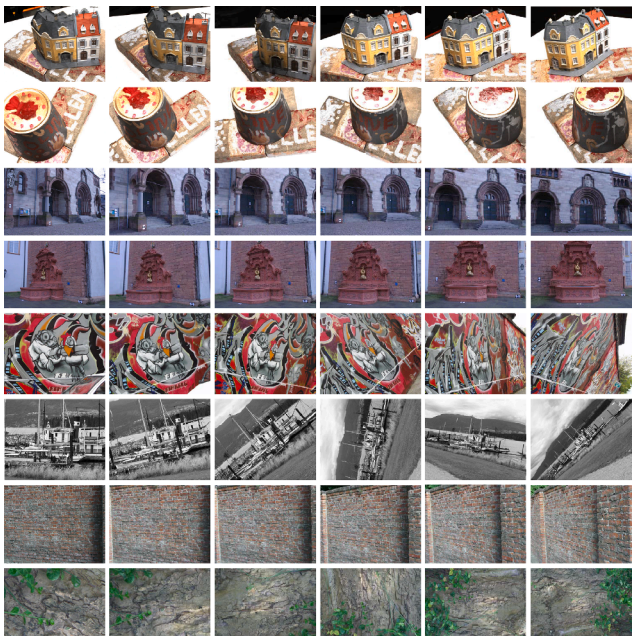


Fig. 10. Different degrees of deformation in 8 scenes. The first two rows are picked from *DTU*, the fourth and fifth rows are selected from *Daisy*, and the remaining four rows are chosen from *VGG*. From left to right, the degree of deformation increases gradually concerning the first column.

function $\mathscr{T}$. A bicubic interpolation algorithm is then used to calculate the intensity at that coordinate in the reference image.

It should be mentioned that in the above pipeline, the recall and precision of a mismatch removal method both significantly affect registration results. Concretely, if the precision is low, the pruned correspondences will contain numerous outliers, which may bias the smooth fitting function generated by TPS and lead to registration failure. If the recall is low, the pruned correspondences may contain too few inliers, which makes that the generated smooth fitting function cannot represent the global transformation well and results in large local fitting errors.

Some intuitive registration results on four typical remote sensing

Table 4
Registration results of ten comparing methods on remote sensing datasets. The average value of RMSE, MAE and MEE are used for evaluation. Red indicates the best, green ranks the second, and blue ranks the third.

Method	RMSE	MAE	MEE
RANSAC	7.458	37.00	5.898
MLESAC	6.987	35.18	5.539
MAGSAC++	7.862	29.45	8.609
CPD	17.63	205.8	2.825
PRGLS	10.25	45.64	10.70
LLT	13.24	53.03	15.14
LPM	9.305	54.74	5.612
GLPM	5.974	46.23	4.233
mTopKRP	7.181	25.19	8.866
NMRC (ours)	5.324	30.64	5.163

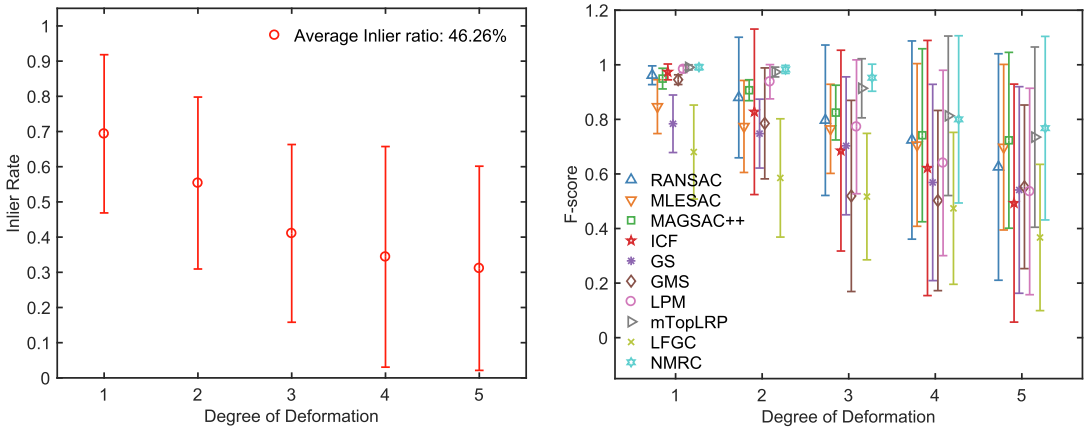


Fig. 11. Robustness test of different methods on an increasing degree of deformation shown in Fig. 10.

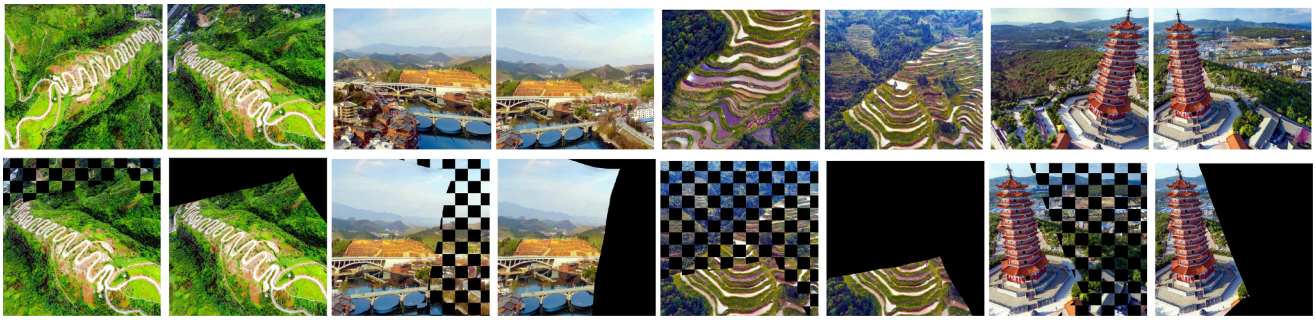


Fig. 12. Qualitative illustration of overall image registration of our NMRC on 4 representative remote sensing image pairs. Top: original input images, where the left and right in each group are sensed and reference images. Bottom: registration results of our NMRC, where the left and right in each group are the 13×13 checkerboard results and the warped sensed images.

Table 5
Details of the experimental datasets on loop closure detection.

Dataset	Description	Image Resolution	#image	Frame Rate (*Hz*)	Rank
KITTI 02	Outdoor, dynamic	1241 × 376	4661	10	Difficult
St1210	Outdoor, dynamic	640 × 480	19251	15	Difficult
Malaga 2009 Parking 6L (Malaga)	Outdoor, slightly dynamic	1024 × 768	3474	7	Difficult

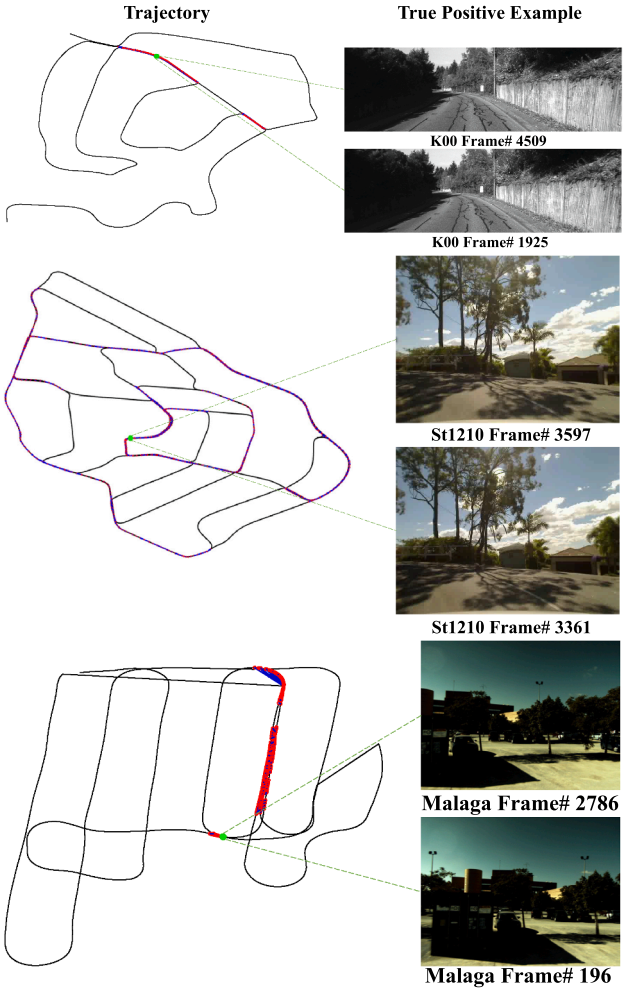


Fig. 13. Illustration of the LCD tasks. From top to bottom: KITTI 02, St1210 and Malaga. From the left to right: robot's trajectory and true positive example. In the left of each row, the black trajectory is obtained from GPS logs, and the results of loop closure identification are acquired under the maximum recall rate at 100% precision. The loop closure pairs are labeled as red hollow points while connecting them using a blue line. The green solid points are the specific illustrations of true-positive detections. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

image pairs are shown in Fig. 12 for visual clarity. The first row indicates the original input images, where the left and right in each group are sensed and reference images, respectively. The second row reports the registration results of NMRC, where the left and right in each group are the checkerboard results and the warped sensed images, respectively. These results indicate that the proposed NMRC can effectively align the overlapping area of all image pairs, especially the fringe areas.

Table 6
Comparative results of the maximum recall rate (%) with 100% accuracy on the three datasets. Red indicates the best, green ranks the second, and blue ranks the third.

Method	KITTI 02	St1210	Malaga
RANSAC (baseline)	78.50	86.22	58.03
MLESAC	76.64	85.85	58.45
MAGSAC++	76.27	86.41	61.89
GMS	79.13	87.89	61.77
LPM	78.82	85.75	61.63
mTopKRP	78.50	84.55	16.62
NMRC (ours)	80.06	87.42	62.05

Six remote sensing datasets are chosen as the test dataset to evaluate the registration performance comprehensively. The first five datasets come from mTopKRP (Jiang et al., 2019), *i.e.*, *UAV*, *SAR*, *PAN*, *CIAP*, and *FE* and undergo projective, similarity or rigid, affine or projective, rigid, and non-rigid, respectively. Moreover, *720 yun* (Liang et al., 2020) is used for non-rigid test. 87 image pairs containing different types of transformation are selected from the abovementioned datasets for evaluation. The average initial correspondence number of these image pairs is approximately 1,080.99, and the inlier rate is only 27.07%. The root mean square error (RMSE), maximum error (MAE), and median error (MEE) are used for measuring the accuracy of image registration with the following definitions:

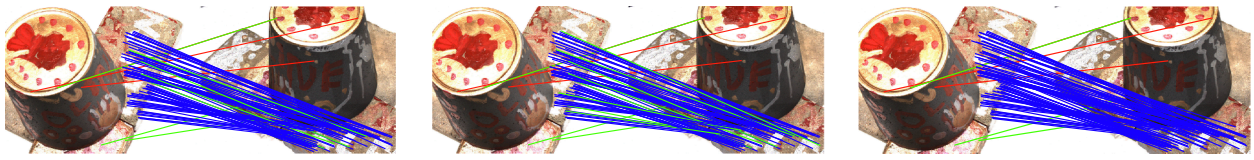
$$RMSE = \sqrt{1 \Big/ L \sum_{i=1}^{L} (r_i^c - \mathscr{T}(s_i^c))^2}, \tag{14}$$

$$MAE = \max\left\{ \sqrt{(r_i^c - \mathscr{T}(s_i^c))^2} \right\}_{i=1}^{L}, \tag{15}$$

$$MEE = median\left\{ \sqrt{(r_i^c - \mathscr{T}(s_i^c))^2} \right\}_{i=1}^{L}, \tag{16}$$

where r_i^c and s_i^c are the corresponding landmarks (*i.e.*, pixel coordinates) of the reference and sensed images respectively, $\mathscr{T}$ is the estimated transformation function from the sensed to the reference image, L denotes the number of selected landmarks, and $\max(\cdot)$ and $median(\cdot)$ return the maximal and median values of a set, respectively. Additionally, the quantitative experiment is manually conducted on the selected landmarks $\{r_i^c, s_i^c\}_{i=1}^{L}$, and the performance evaluation is measured by calculating the RMSE, MAE, and MEE of 20 pairs of landmarks distributed in easily identifiable locations around the region of interest. The statistical results on 87 selected image pairs are reported in Table 4[1]. Nine representative methods are used for comparison, including RANSAC (Fischler and Bolles, 1981), MLESAC (Torr and Zisserman, 2000), MAGSAC++ (Barath et al., 2021), CPD (Myronenko and Song, 2010), PRGLS (Ma et al., 2016), LLT (Ma et al., 2015), LPM (Ma et al., 2019), GLPM (Ma et al., 2018), and mTopKRP (Jiang et al., 2019). The table reveals that NMRC can achieve the best performance of RMSE, the second-best performance of MAE, and the third-best performance of MEE. CPD obtains the best MEE performance followed by GLPM and NMRC. However, CPD is insufficiently robust to cope with image registration tasks due to the poor metrics of RMSE and MAE. RANSAC, MLESAC and MAGSAC++ achieve stable performance owing to their global geometrical constraints but fail to address the non-rigid deformations. The relatively poor performance of LLT and PRGLS is mainly caused by the low putative inlier rate. In addition, LPM performs poorly

[1] Due to the existence of false correspondences, it may lead to a large local fitting error, which will cause a high standard deviation.



(a) LPM (b) mTopKRP (c) NMRC

Fig. 14. A failure case where the relaxation (*e.g.*, neighborhood-aware) approaches such as LPM, mTopKRP and the proposed NMRC fail to distinguish inliers and outliers in a fairly sparse area (*i.e.*, the local area containing false negatives in the result of NMRC).

for the same reason, *i.e.*, some gross outliers may be preserved due to its not strict measuring criterion. mTopKRP has the best performance of MAE and performs relatively well considering RMSE due to the K-NN ranking list metric. Meanwhile, GLPM also has a rather promising performance because of its guided matching strategy. The proposed NMRC approach can realize the most satisfying performance.

4.4. Results on loop closure detection

Loop closure detection (LCD), which entails the correct identification of previously visited areas in the environment, is an essential component in Visual SLAM systems (Zhang et al., 2021). Existing LCD methods typically apply RANSAC to verify the loop closing pairs due to its capability to establish reliable correspondences and estimate transformation between image pairs simultaneously. However, RANSAC hinges on a pre-defined parametric model and cannot tackle complex transforms, *e.g.*, non-rigid deformation.

Consequently, RANSAC is used as the baseline to substantiate the validity of the NMRC used to verify candidate frames for LCD tasks. In particular, the rather prevalent bag-of-words (BoW) is utilized to select loop closure candidates to decrease the computational requirement and then adopt mismatch removal methods to establish the image correspondences to detect the loop closing pairs. Three publicly available datasets (Zhang et al., 2021), *i.e.*, KITTI 02, St1210, and Malaga 2009 Parking 6L (Malaga), which are summarized in Table 5, are selected to provide a comprehensive evaluation on this task.

4.4.1. Qualitative illustration

First, some qualitative results on the LCD performance of NMRC are presented for visual clarity. According to the data recorded by odometry, the trajectories of robots on each dataset are drawn in the black lines of Fig. 13. The loop closure pairs are labeled as red hollow points while connecting them with blue lines. The green solid points are the specific illustrations of true-positive detection.

4.4.2. Quantitative comparison

The loop closure verified by the algorithm in an LCD system provides imprecise information, triggering the inevitable performance degradation of the entire system, especially when the LCD module considers a false loop closing pair as true. Meanwhile, an ideal LCD module should have a high recall, *i.e.*, detecting as many loop closing pairs as possible. Accordingly, the maximum recall rate is a significant evaluation index of the performance of an algorithm when the precision is 100%.

The comparative results are summarized in Table 6. Six mismatch removal methods, such as RANSAC, MLESAC, MAGSAC++, GMS, LPM, and mTopKRP, are used for comparison. The table shows that NMRC can realize the highest recall rates for 100% precision on two of the three datasets, *i.e.*, KITTI 02 and Malaga. Although NMRC performs slightly poorly on St1210 compared with GMS, the proposed method still ranks second and achieves a rather satisfying level. Overall, NMRC can achieve the most stable performance on different datasets.

5. Conclusion

A robust feature matching approach named NMRC, which is based on the stable local topology of the potential true matches between two feature sets, is proposed in this paper. The neighborhood representations of each putative correspondence are obtained based on manifold learning, and the consensus of neighborhood manifold representation is then measured using the square of L_2 norm. Particularly, a neighborhood similarity-based iterative filtering strategy is devised to boost the robustness under the circumstance of seriously deteriorated data. Meanwhile, the idea is formulated into a mathematical model, and a closed-form solution is derived with linearithmic time complexity. The experimental results on general feature matching and two real-world tasks demonstrate that the proposed strategy outperforms the state-of-the-art competitors.

Notably, although the neighborhood manifold representation is highly effective, such a representation relies on a good initialization to capture sufficient reliable correspondences in a local area. This will support to construct more accurate neighborhood manifold representation. Fortunately, it works well for most cases under the proposed iterative filtering strategy. However, this strategy would be limited when the putative correspondences are pretty sparse in a local area, which may cause difficulties in representing the neighborhood manifold. Note that this is a common issue for relaxation methods, such as LPM (Ma et al., 2019) and mTopKRP (Jiang et al., 2019), which are derived from the local consensus assumption. A failure case is presented in Fig. 14, which indicates that LPM, mTopKRP and the proposed NMRC fail to distinguish inliers and outliers in the local area with extremely sparse putative correspondences. Nevertheless, the proposed NMRC still has a better overall performance. Future research will focus on designing a better feature matcher to establish more valid putative correspondences and fully utilizing the local and global geometric information to deal with this situation.

CRediT authorship contribution statement

Jiayi Ma: Conceptualization, Methodology. **Zizhuo Li:** Methodology. **Kaining Zhang:** Methodology. **Zhenfeng Shao:** Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was sponsored by the National Natural Science Foundation of China (61773295).

References

Barath, D., Noskova, J., Matas, J., 2021. Marginalizing sample consensus. IEEE Trans. Pattern Anal. Mach. Intell.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: Speeded up robust features. In: Proceedings of the European Conference on Computer Vision, pp. 404–417.

Belongie, S., Malik, J., Puzicha, J., 2002. Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell. 24, 509–522.

Bentley, J.L., 1975. Multidimensional binary search trees used for associative searching. Commun. ACM 18, 509–517.

Besl, P., McKay, N., 1992. A method for registration of 3-d shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14, 239–256.
Bian, J.W., Lin, W.Y., Liu, Y., Zhang, L., Yeung, S.K., Cheng, M.M., Reid, I., 2020. Gms: Grid-based motion statistics for fast, ultra-robust feature correspondence. Int. J. Comput. Vision 128, 1580–1593.
Chen, J., Chen, S., Chen, X., Yang, Y., Xing, L., Fan, X., Rao, Y., 2021. Lsv-anet: Deep learning on local structure visualization for feature matching. In: IEEE Trans. Geosci. Remote Sens..
Chui, H., Rangarajan, A., 2003. A new point matching algorithm for non-rigid registration. Comput. Vis. Image Underst. 89, 114–141.
DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. Superpoint: Self-supervised interest point detection and description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 224–236.
Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24, 381–395.
Huber, P.J., 1981. Robust statistics. John Wiley & Sons.
Jiang, X., Jiang, J., Fan, A., Wang, Z., Ma, J., 2019. Multiscale locality and rank preservation for robust feature matching of remote sensing images. IEEE Trans. Geosci. Remote Sens. 57, 6462–6472.
Jiang, X., Ma, J., Xiao, G., Shao, Z., Guo, X., 2021. A review of multimodal image matching: Methods and applications. Inform. Fusion 73, 22–71.
Leordeanu, M., Hebert, M., 2005. A spectral technique for correspondence problems using pairwise constraints. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1482–1489.
Li, X., Hu, Z., 2010. Rejecting mismatches by correspondence function. Int. J. Comput. Vision 89, 1–17.
Liang, L., Zhao, W., Hao, X., Yang, Y., Yang, K., Liang, L., Yang, Q., 2020. Image registration using two-layer cascade reciprocal pipeline and context-aware dissimilarity measure. Neurocomputing 371, 1–14.
Li, J., Hu, Q., Ai, M., 2019. Lam: Locality affine-invariant feature matching. ISPRS J. Photogram. Remote Sens. 154, 28–40.
Li, J., Hu, Q., Ai, M., 2020. Gesac: Robust graph enhanced sample consensus for point cloud registration. ISPRS J. Photogram. Remote Sens. 167, 363–374.
Li, J., Hu, Q., Ai, M., 2020. Rift: Multi-modal image matching based on radiation-variation insensitive feature transform. IEEE Trans. Image Process. 29, 3296–3310.
Li, J., Zhao, P., Hu, Q., Ai, M., 2020. Robust point cloud registration based on topological graph and cauchy weighted lq-norm. ISPRS J. Photogram. Remote Sens. 160, 244–259.
Liu, H., Yan, S., 2010. Common visual pattern discovery via spatially coherent correspondences. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1609–1616.
Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60, 91–110.
Ma, J., Zhao, J., Tian, J., Yuille, A.L., Tu, Z., 2014. Robust point matching via vector field consensus. IEEE Trans. Image Process. 23, 1706–1721.
Ma, J., Zhou, H., Zhao, J., Gao, Y., Jiang, J., Tian, J., 2015. Robust feature matching for remote sensing image registration via locally linear transforming. IEEE Trans. Geosci. Remote Sens. 53, 6469–6481.
Ma, J., Zhao, J., Yuille, A.L., 2016. Non-rigid point set registration by preserving global and local structures. IEEE Trans. Image Process. 25, 53–64.
Ma, J., Jiang, J., Zhou, H., Zhao, J., Guo, X., 2018. Guided locality preserving feature matching for remote sensing image registration. IEEE Trans. Geosci. Remote Sens. 56, 4435–4447.
Ma, J., Jiang, X., Jiang, J., Zhao, J., Guo, X., 2019. Lmr: Learning a two-class classifier for mismatch removal. IEEE Trans. Image Process. 28, 4045–4059.
Ma, J., Zhao, J., Jiang, J., Zhou, H., Guo, X., 2019. Locality preserving matching. Int. J. Comput. Vision 127, 512–531.
Ma, J., Jiang, X., Fan, A., Jiang, J., Yan, J., 2021. Image matching from handcrafted to deep features: A survey. Int. J. Comput. Vision 129, 23–79.
Maier, J., Humenberger, M., Murschitz, M., Zendel, O., Vincze, M., 2016. Guided matching based on statistical optical flow for fast and robust correspondence analysis. In: Proceedings of the European Conference on Computer Vision, pp. 101–117.
Moo Yi, K., Trulls, E., Ono, Y., Lepetit, V., Salzmann, M., Fua, P., 2018. Learning to find good correspondences. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2666–2674.
Myronenko, A., Song, X., 2010. Point set registration: Coherent point drift. IEEE Trans. Pattern Anal. Mach. Intell. 32, 2262–2275.
Rousseeuw, P.J., Leroy, A.M., 2005. Robust regression and outlier detection, vol. 589. John wiley & sons.
Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326.
Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. Orb: An efficient alternative to sift or surf. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2564–2571.
Sarlin, P.E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020. Superglue: Learning feature matching with graph neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4938–4947.
Schonberger, J.L., Frahm, J.M., 2016. Structure-from-motion revisited, in. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4104–4113.
Shao, F., Liu, Z., An, J., 2020. A discriminative point matching algorithm based on local structure consensus constraint. IEEE Geosci. Remote Sens. Lett. 18, 1366–1370.
Shao, F., Liu, Z., An, J., 2021. Feature matching based on minimum relative motion entropy for image registration. IEEE Trans. Geosci. Remote Sens.
Simonovsky, M., Gutiérrez-Becker, B., Mateus, D., Navab, N., Komodakis, N., 2016. A deep metric for multimodal registration. In: Proceedings of the International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 10–18.
Sun, K., Tao, W., Qian, Y., 2020a. Guide to match: Multi-layer feature matching with a hybrid gaussian mixture model. IEEE Trans. Multimedia 22, 2246–2261.
Sun, W., Jiang, W., Trulls, E., Tagliasacchi, A., Yi, K.M., 2020b. Acne: Attentive context normalization for robust permutation-equivariant learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 11286–11295.
Torr, P.H., Zisserman, A., 2000. Mlesac: A new robust estimator with application to estimating image geometry. Comput. Vis. Image Underst. 78, 138–156.
Vedaldi, A., Fulkerson, B., 2010. Vlfeat: An open and portable library of computer vision algorithms. In: Proceedings of the ACM International Conference on Multimedia, pp. 1469–1472.
Wang, C., Wang, L., Liu, L., 2014. Progressive mode-seeking on graphs for sparse feature matching. In: Proceedings of the European Conference on Computer Vision, pp. 788–802.
Wong, H.S., Chin, T.J., Yu, J., Suter, D., 2011. Dynamic and hierarchical multi-structure geometric model fitting. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1044–1051.
Yang, Y., Ong, S.H., Foong, K.W.C., 2015. A robust global and local mixture distance based non-rigid point set registration. Pattern Recogn. 48, 156–173.
Yi, K.M., Trulls, E., Lepetit, V., Fua, P., 2016. Lift: Learned invariant feature transform. In: Proceedings of the European Conference on Computer Vision, pp. 467–483.
Yu, Q., Ni, D., Jiang, Y., Yan, Y., An, J., Sun, T., 2021. Universal sar and optical image registration via a novel sift framework based on nonlinear diffusion and a polar spatial-frequency descriptor. ISPRS J. Photogram. Remote Sens. 171, 1–17.
Zhang, H., Ma, J., 2021. Gtp-pnet: A residual learning network based on gradient transformation prior for pansharpening. ISPRS J. Photogram. Remote Sens. 172, 223–239.
Zhang, J., Sun, D., Luo, Z., Yao, A., Zhou, L., Shen, T., Chen, Y., Quan, L., Liao, H., 2019. Learning two-view correspondences and geometry using order-aware network. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5845–5854.
Zhang, K., Jiang, X., Ma, J., 2021. Appearance-based loop closure detection via locality-driven accurate motion field learning. IEEE Trans. Intel. Transp. Syst.
Zitova, B., Flusser, J., 2003. Image registration methods: a survey. Image Vis. Comput. 21, 977–1000.