



Robust feature matching via advanced neighborhood topology consensus

Yizhang Liu^{a,b}, Yanping Li^c, Luanyuan Dai^a, Changcai Yang^{a,*}, Lifang Wei^a, Taotao Lai^d, Riqing Chen^a

^a Digital Fujian Research Institute of Big Data for Agriculture and Forestry, College of Computer and Information Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China

^b School of Software Engineering, Tongji University, Shanghai 201804, China

^c Department of Computer Science and Technology, Tongji University, Shanghai 201804, China

^d College of Computer and Control Engineering, Minjiang University, Fuzhou 350108, China

ARTICLE INFO

Article history:

Received 15 August 2020

Revised 21 September 2020

Accepted 27 September 2020

Available online 6 October 2020

Communicated by Steven Hoi

Keywords:

Feature matching

Outlier removal

Guided matching strategy

Neighborhood topology

ABSTRACT

Feature matching is one of the key techniques in many vision-based tasks, which aims to establish reliable correspondences between two sets of features. In this paper, we present a new feature matching method, which formulates the matching of two feature sets as a mathematical model based on two common consistency constraints. We first propose an advanced consensus of neighborhood topology, which can better exploit the consensus of topological structures to identify inliers. In order to have reliable neighborhood information for the feature points, a subset with high percentage inliers obtained by a guided matching strategy from the putative matches for the neighborhood construction is used. We demonstrate the advantages of our proposed method on various real image pairs. The results demonstrate that the proposed method is superior to the state-of-the-art feature matching methods.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Feature matching is the fundamental prerequisite for a lot of computer vision techniques. Many tasks in computer vision, such as image registration [1–3], image fusion [4,5], 3D reconstruction [6,7], point set registration [8,9], can be typically considered as feature matching problem. The features to be matched are usually extracted from images of similar or the same scene, which may be obtained at different times, from different viewpoints, or by different sensors. The goal of feature matching is to establish reliable correspondence between the two images. The feature matching problem is usually tackled by a pipeline, i.e., first constructing initial correspondences by some feature extraction and description techniques (e.g., SIFT [10]) and then selecting reliable correspondences from the putative sets with some outlier removal algorithms. Due to ambiguities of the descriptors, the putative sets usually contain a number of outliers. Therefore, a robust outlier removal algorithm is particularly necessary for some feature matching task [11,12].

In order to match two feature sets accurately, many approaches have been developed in the past few decades. Feature matching

methods can be roughly divided into four categories, i.e., resampling methods, non-parametric interpolation methods, graph matching methods, and learning-based methods. The resampling methods [13–16] adopt a hypothesis and verification procedure, sampling the minimal subset from all inliers to calculate parameters of the model through resampling. Nevertheless, to find the minimal subset from all inliers, the iteration number needs to be exponentially increased in cases of high outlier percentages in the putative set. Moreover, a predefined parametric model is less efficient to deal with image pairs that undergo complex non-rigid transformations. A number of non-parametric interpolation methods [17–21] have been proposed to deal with the above mentioned problems. These methods usually make some prior assumption such as geometric consistency between inliers to learn a pair of correspondence functions or interpolate a motion field, where the correspondences are slow-and-smooth. Compared with resampling methods, the non-parametric interpolation methods can deal with non-rigid transformations, whereas they commonly have cubic complexities which restrains their applications on real-time tasks. Graph matching methods [22–25] represent the matching problem as an integer quadratic programming (IQP) problem that forces the geometry between two feature sets to be preserved, which does not rely on a specific transformation. The optimal solution of it determines the final matching results. However, these methods only consider the spatial information and ignore the local

* Corresponding author.

E-mail addresses: lyz8023lyp@gmail.com (Y. Liu), changcaiyang@gmail.com (C. Yang), riqing.chen@fafu.edu.cn (R. Chen).

descriptor information, which are not robust enough to deal with different image deformations, meanwhile, the non-polynomial-hard nature will also result in large computational complexity, which is not suitable for large-scale matching problems. Learning-based methods [26–30] are another technique to deal with the feature matching problem. Although it has been verified that the representations obtained by those using deep learning architectures are better than hand-crafted representations, the putative set still contains a large number of mismatches. Therefore, a mismatch removal method remains necessary. Besides, the GANs have been used to the point set registration, which regard the point clouds as the probability distributions. The aim of these methods is to learn a critic network confusing the source point sets and the target point sets. Some of them just need the spatial information of the points and do not construct the initial correspondences between the two point sets. Therefore, they are not suitable for the image feature matching problem.

Most recently, several coherence constraint and local neighborhood consistency based methods are proposed such as grid-based motion statistics (GMS) [31], coherence-based decision boundaries (CODE) [6], and locality preserving matching (LPM) [32], which achieve promising results. In LPM, it is based on the observation that the absolute distance between the corresponding feature points may vary a lot when the image pairs degrade due to complex deformations, but in the local neighborhood of the feature points, the local neighborhood structure and the topological structures are generally well preserved due to physical constraints. The advantage of LPM is that the simple formulation makes it possible to remove mismatches from thousands of putative matches within a few milliseconds. Nevertheless, the main deficiencies of LPM are that the consensus of neighborhood topology cannot distinguish inliers from the putative sets well and it is apt to have an over-penalization on outliers. Therefore, as the outlier ratio of the putative sets increases, its neighborhood construction will become unreliable. In this paper, we propose a new robust feature matching method based on the advanced neighborhood topology consensus (ANTC), which can efficiently distinguish inliers from the whole putative sets. The main contributions of this work are twofold. First, we propose an advanced consensus of neighborhood topology, which can expand the distributions between outliers and inliers and better exploit the consensus of topological structures to identify inliers. Second, to deal with high ratio outliers in the putative set, a guided matching strategy is used to establish a subset with high ratio of inliers, which enables the neighborhood construction based on this subset more reliable leading to an almost perfect matching result. Extensive experiments on three public available datasets referring to different image deformation (e.g. rigid transformation, wide baseline, repetitive structures, non-rigid transformation and so on) compared with the state-of-the-art methods show that our proposed ANTC are better in terms of precision, recall, and F-score.

The remainder of this paper is organized as follows. In Section 2, we first review the original LPM and then introduce our proposed ANTC in detail, involving the advanced consensus of neighborhood topology and the guided matching strategy. The image feature matching experimental results on different types of image datasets are shown in Section 3, and the conclusions are drawn in Section 4.

2. Method

In this section, we first review the locality preserving matching in Section 2.1. Then, we introduce two novel components of the proposed ANTC: the advanced consensus of neighborhood topology in Section 2.2 and the guided matching strategy in Section 2.3.

In Section 2.4 we present the proposed complete ANTC method. In Section 2.5 we discuss the computational complexity of the proposed method.

2.1. Locality preserving matching

In this study, the input data are N putative matches $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$ extracted from two input images by using the well-known feature descriptor (SIFT [10]), where \mathbf{x}_i and \mathbf{y}_i are the spatial coordinates of corresponding features. The aim of this paper is to remove mismatches in S to establish accurate correspondences. Ma et al. [32] has shown that mismatch removal problem could be solved by seeking the minimal cost function:

$$\mathcal{C}(\mathcal{I}; S, \lambda) = \sum_{i \in \mathcal{I}} \frac{1}{2K} \left(\sum_{j|\mathbf{x}_j \in N_{\mathbf{x}_i}} (d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{y}_i, \mathbf{y}_j))^2 + \sum_{j|\mathbf{y}_j \in N_{\mathbf{y}_i}} (d(\mathbf{x}_i, \mathbf{x}_j) - d(\mathbf{y}_i, \mathbf{y}_j))^2 \right) + \lambda(N - |\mathcal{I}|), \quad (1)$$

where \mathcal{I} is the unknown inlier set, $N_{\mathbf{x}}$ is the neighborhood of feature point \mathbf{x} consisting of K nearest neighbors of the point \mathbf{x} , d is a certain distance metric (e.g. Euclidean distance), and $|\cdot|$ denotes the cardinality of a set. The parameter $\lambda > 0$ controls the tradeoff between the first and the second items. According to the motion coherence theory [33], if two features $\mathbf{x}_i, \mathbf{x}_j$ in one image are close, they probably belongs to the same object and tend to move together. The description above can be expressed by the formulation, $d(\mathbf{x}_i, \mathbf{x}_j) \approx d(\mathbf{y}_i, \mathbf{y}_j)$. We use K nearest neighbor to describe each feature's neighborhood and then the first term of the cost function is close to zero for an inlier and is much larger than zero for an outlier. Besides, in order to be scale-invariant, the distance has been quantized as follow:

$$d(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 0, & \mathbf{x}_j \in \mathcal{N}_{\mathbf{x}_i} \\ 1, & \mathbf{x}_j \notin \mathcal{N}_{\mathbf{x}_i} \end{cases}, \quad d(\mathbf{y}_i, \mathbf{y}_j) = \begin{cases} 0, & \mathbf{y}_j \in \mathcal{N}_{\mathbf{y}_i} \\ 1, & \mathbf{y}_j \notin \mathcal{N}_{\mathbf{y}_i} \end{cases} \quad (2)$$

The cost function described above only preserves the intersection of neighbors and ignores their topological structures, which will be problematic sometimes. Here, we give an example in Fig. 1. Both the inlier $(\mathbf{x}_i, \mathbf{y}_i)$ in Fig. 1(a) and the outlier in Fig. 1(b) will have the same contribution to the cost function above and they could not be told apart. Thus, the consensus of neighborhood topology which consists of the angle and the ratio of length between \mathbf{v}_i and \mathbf{v}_j can be expressed as follows:

$$s(\mathbf{v}_i, \mathbf{v}_j) = \frac{\min\{|\mathbf{v}_i|, |\mathbf{v}_j|\}}{\max\{|\mathbf{v}_i|, |\mathbf{v}_j|\}} \cdot \frac{(\mathbf{v}_i, \mathbf{v}_j)}{|\mathbf{v}_i| \cdot |\mathbf{v}_j|} \quad (3)$$

where \mathbf{v}_i and \mathbf{v}_j are the displacement vectors of putative match $(\mathbf{x}_i, \mathbf{y}_i)$ and one of its neighboring matches $(\mathbf{x}_j, \mathbf{y}_j)$, respectively. More specifically, $\mathbf{v}_i = \mathbf{y}_i - \mathbf{x}_i$ and $\mathbf{v}_j = \mathbf{y}_j - \mathbf{x}_j$. (\cdot, \cdot) denotes the inner product. Obviously, the consensus of neighborhood topology $s(\mathbf{v}_i, \mathbf{v}_j) \in [-1, 1]$. The larger the value of $s(\mathbf{v}_i, \mathbf{v}_j)$, the higher the consensus of neighborhood topology can be obtained. A predefined threshold τ was used to distinguish the inliers from the putative set as follows:

$$\hat{d}(\mathbf{v}_i, \mathbf{v}_j) = \begin{cases} 0, & s(\mathbf{v}_i, \mathbf{v}_j) \geq \tau \\ 1, & s(\mathbf{v}_i, \mathbf{v}_j) < \tau \end{cases} \quad (4)$$

Besides, an $N \times 1$ binary vector \mathbf{p} , where $p_i \in \{0, 1\}$, was introduced to indicate whether the i -th correspondence $(\mathbf{x}_i, \mathbf{y}_i)$ in putative set S is correct. Specifically, $p_i = 1$ denotes inlier and $p_i = 0$ points to outlier. Thus, the objective function in Eq. (1) can be written as the following minimization problem:

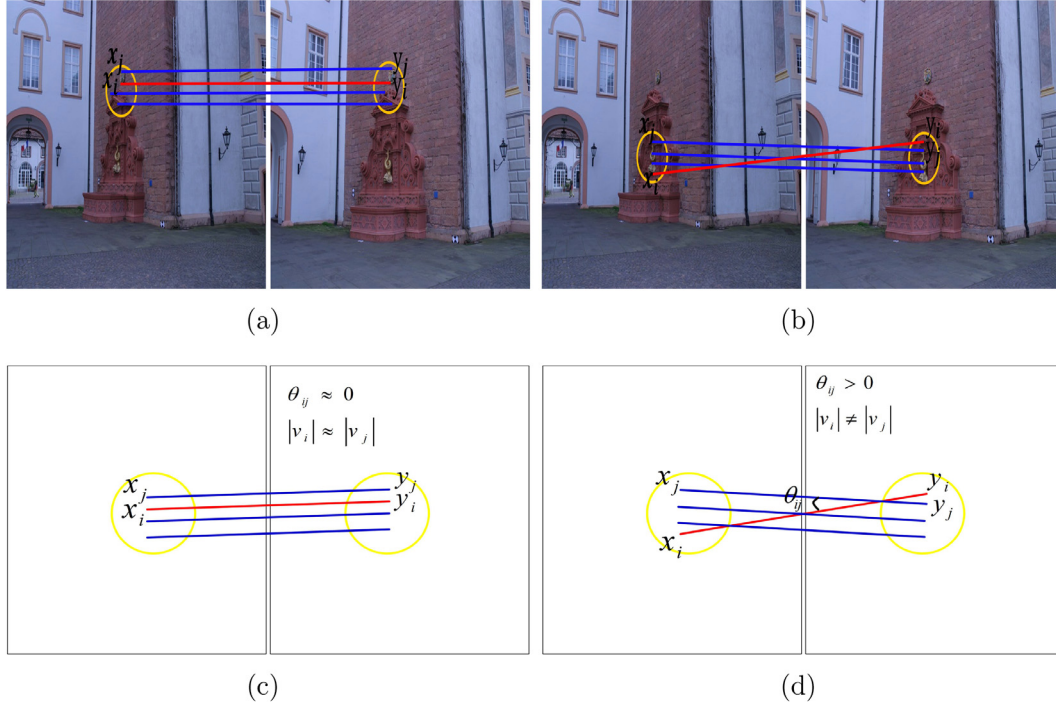


Fig. 1. Local neighborhood topology. (a) An inlier ($\mathbf{x}_i, \mathbf{y}_i$) and its neighboring putative matches. (b) An outlier ($\mathbf{x}_i, \mathbf{y}_i$) and its neighboring putative matches. (c) Local neighborhood topology corresponding to (a). (d) Local neighborhood topology corresponding to (b). For an inlier ($\mathbf{x}_i, \mathbf{y}_i$), the value of angle θ_{ij} is close to zero and the length $|v_i|$ is similar to the length $|v_j|$ corresponding to a putative match ($\mathbf{x}_j, \mathbf{y}_j$) in its local neighborhood. For an outlier ($\mathbf{x}_i, \mathbf{y}_i$), the value of angle θ_{ij} is larger than zero and the length $|v_i|$ differ greatly from the length $|v_j|$ corresponding to a putative match ($\mathbf{x}_j, \mathbf{y}_j$) in its local neighborhood.

$$\mathbf{C}(\mathbf{p}; S, \lambda, \tau) = \sum_{i=1}^N \frac{p_i}{K} \left(\sum_{j|\mathbf{x}_j \in N_{\mathbf{x}_i}} d(\mathbf{y}_i, \mathbf{y}_j) + \sum_{j|\mathbf{x}_j \in N_{\mathbf{x}_i}, \mathbf{y}_j \in N_{\mathbf{y}_i}} \hat{d}(\mathbf{v}_i, \mathbf{v}_j) \right) + \lambda \left(N - \sum_{i=1}^N p_i \right) \quad (5)$$

2.2. Advanced consensus of neighborhood topology

We follow the similar idea of the LPM [32], but, rather than introducing a single model for the local neighborhood topology, we model the consensus of local neighborhood topology as a soft decision. Fig. 1 shows the local neighborhood topologies of inlier and outlier. $\theta_{ij} = \langle \mathbf{v}_i, \mathbf{v}_j \rangle \in [0, \pi]$ characterizes the angle between \mathbf{v}_i and \mathbf{v}_j . From Fig. 1(a) and (c), we can see that the value of angle θ_{ij} is close to zero and the length $|v_i|$ corresponding to the inlier ($\mathbf{x}_i, \mathbf{y}_i$) is similar to the length $|v_j|$ corresponding to a certain putative match ($\mathbf{x}_j, \mathbf{y}_j$) in its local neighborhood. From Fig. 1(b) and (d), we can see that the value of angle θ_{ij} is larger than zero and the length $|v_i|$ for the outlier ($\mathbf{x}_i, \mathbf{y}_i$) differ greatly from the length $|v_j|$ corresponding to a certain putative match ($\mathbf{x}_j, \mathbf{y}_j$) in its local neighborhood. That is to say, both length and angle of the displacement vector corresponding to an inlier differ little from the putative matches in its local neighborhood, while both length and angle of the displacement vector corresponding to an outlier differ greatly from the putative matches in its local neighborhood. Based on the above analysis, we define the advanced consensus of neighborhood topology as a soft decision, which can be written as

$$c^a(\mathbf{v}_i, \mathbf{v}_j) = \frac{1}{\hat{\sigma}} \exp \left\{ -\frac{(R_{ij} + \xi \hat{\theta}_{ij})^2}{2\hat{\sigma}^2} \right\} \quad (6)$$

where $R_{ij} = \frac{\max\{|\mathbf{v}_i|, |\mathbf{v}_j|\}}{\min\{|\mathbf{v}_i|, |\mathbf{v}_j|\}} - 1 \geq 0$ represents the length ratio between \mathbf{v}_i and \mathbf{v}_j . ξ is a weight that controls the relative influence of the ratio of length and angle between \mathbf{v}_i and \mathbf{v}_j . $\hat{\sigma}$ is the inlier noise scale of the putative set. Note that, for an inlier, it will have great consistency with its neighboring matches both in length and angle, resulting in a large value of $c^a(\mathbf{v}_i, \mathbf{v}_j)$ and vice versa.

In Eq. (6), \mathbf{v}_i should compare with all its neighboring displacement vectors of putative matches, which will cause more time costs. Similar to LMR [27], we choose to calculate the average displacement vector of the n_a neighboring putative matches i.e. $\bar{\mathbf{v}}_j$. By comparing the difference between \mathbf{v}_i and $\bar{\mathbf{v}}_j$, the consensus of neighborhood topology can be exploited. More specifically, changes in the topological structures of the n_a elements relative to \mathbf{x}_i and \mathbf{y}_j will cause significant differences in both length and angle between \mathbf{v}_i and $\bar{\mathbf{v}}_j$. The advanced consensus of neighborhood topology in Eq. (6) can be rewritten as follows:

$$\tilde{c}^a(\mathbf{v}_i, \bar{\mathbf{v}}_j) = \frac{1}{\hat{\sigma}} \exp \left\{ -\frac{(\tilde{R}_i + \xi \tilde{\theta}_i)^2}{2\hat{\sigma}^2} \right\} \quad (7)$$

where $\tilde{R}_i = \frac{\max\{|\mathbf{v}_i|, |\bar{\mathbf{v}}_j|\}}{\min\{|\mathbf{v}_i|, |\bar{\mathbf{v}}_j|\}} - 1 \geq 0$ is the ratio of length between \mathbf{v}_i and $\bar{\mathbf{v}}_j$. $\tilde{\theta}_i = \langle \mathbf{v}_i, \bar{\mathbf{v}}_j \rangle \in [0, \pi]$ is the angle between \mathbf{v}_i and $\bar{\mathbf{v}}_j$. The advanced consensus of neighborhood topology can robustly recover the inliers from the putative set.

Due to advanced consensus of neighborhood topology, the range of value of $\tilde{c}^a(\mathbf{v}_i, \bar{\mathbf{v}}_j)$ is enlarged and it is beneficial to use a pre-defined threshold to distinguish the inliers from the putative set. In order to adapt to different scales of image matching, we use a quantized distance metric between \mathbf{v}_i and $\bar{\mathbf{v}}_j$ with a precalculated threshold τ :

$$\hat{d}(\mathbf{v}_i, \tilde{\mathbf{v}}_j) = \begin{cases} -1, & \tilde{c}^a(\mathbf{v}_i, \tilde{\mathbf{v}}_j) \geq \tau \\ 1, & \tilde{c}^a(\mathbf{v}_i, \tilde{\mathbf{v}}_j) < \tau \end{cases} \quad (8)$$

with τ :

$$\tau = \frac{1}{\hat{\sigma}} \exp \left\{ -\frac{(R_t + \xi \theta_t)^2}{2\hat{\sigma}^2} \right\} \quad (9)$$

where R_t and θ_t are the threshold values of ratio of length and angle between \mathbf{v}_i and $\tilde{\mathbf{v}}_j$, respectively. Specifically, when the values of \tilde{R}_i and $\tilde{\theta}_i$ are both smaller than R_t and θ_t , respectively, we consider that \mathbf{v}_i and $\tilde{\mathbf{v}}_j$ have high consensus of neighborhood topology, which is in accord with the intuitive observation. In the case of the above distance, considering that the optimal value of K may vary with different image data, a multi-scale neighborhood construction strategy is designed with sizes $K = \{K_m\}_{m=1}^M$, e.g., $\{N_{\mathbf{x}_i}^{K_m}\}_{m=1}^M$ and $\{N_{\mathbf{y}_i}^{K_m}\}_{m=1}^M$, where $\{N_{\mathbf{x}_i}^{K_m}\}$ is the neighborhood of point \mathbf{x}_i formed by its K_m nearest neighbors under Euclidean distance. We empirically set $\hat{\sigma}$ to be 0.5 throughout our experiments. $\hat{d}(\mathbf{v}_i, \tilde{\mathbf{v}}_j)$ in Eq. (8) is different from that in LPM. Specifically, when a putative match is not consistent with its neighboring matches, we make the cost function decrease instead of being invariant. In this case, the correctness of a putative match will not be totally affected by either the consensus of neighborhood elements or the consensus of neighborhood topology.

Following [27], we then solve the removing mismatches problem by the following objective function composed of the advanced consensus of neighborhood topology:

$$\begin{aligned} \mathcal{Q}(\mathbf{p}; S, \lambda, \tau, \xi) = & \sum_{i=1}^N \frac{p_i}{M} \sum_{m=1}^M \frac{1}{K_m} \left(\sum_{j|\mathbf{x}_j \in N_{\mathbf{x}_i}^{K_m}} d(\mathbf{y}_i, \mathbf{y}_j) + K_m \cdot \hat{d}(\mathbf{v}_i, \tilde{\mathbf{v}}_j) \right) \\ & + \lambda \left(N - \sum_{i=1}^N p_i \right). \end{aligned} \quad (10)$$

where $1/M$ is used to normalize the influence of different scales of neighborhood. The first term consists of two constrains, namely, the consensus of neighborhood elements and the advanced consensus of neighborhood topology. The second term is a regularization term, which is used to discourage the outliers. The accurate correspondences can be obtained by minimizing Eq. (10).

If we reorganize Eq. (10) by merging the terms with respect to p_i , the objective function (10) can be written in the form

$$\mathcal{Q}(\mathbf{p}; S, \lambda, \tau, \xi) = \sum_{i=1}^N p_i (\tilde{d}_i - \lambda) + \lambda N \quad (11)$$

where

$$\tilde{d}_i = \sum_{m=1}^M \frac{1}{MK_m} \left(\sum_{j|\mathbf{x}_j \in N_{\mathbf{x}_i}^{K_m}} d(\mathbf{y}_i, \mathbf{y}_j) + K_m \cdot \hat{d}(\mathbf{v}_i, \tilde{\mathbf{v}}_j) \right) \quad (12)$$

indicates the degree that the i -th putative match $(\mathbf{x}_i, \mathbf{y}_i)$ meets the geometric consistency of preserving the locality and topology nature. Specifically, a true match will lead to small cost, whereas the mismatch will bring large cost.

Since the neighborhood relationship with respect to a given putative set is fixed, the values $\{\tilde{d}_i\}_{i=1}^N$ can be pre-calculated. Therefore, the only unknown variable of the cost function in Eq. (11) is p_i . We use a predefined threshold to distinguish the outliers and inliers. Specifically, the matches with cost larger than the threshold λ will increase the cost function. Conversely, the matches with cost smaller than the threshold λ will decrease the cost function. Thus, the optimal solution of p can be determined as follows:

$$p_i = \begin{cases} 1, & \tilde{d}_i \leq \lambda \\ 0, & \tilde{d}_i > \lambda \end{cases}, \quad i = 1, \dots, N. \quad (13)$$

Then, the optimal inlier set \mathcal{I} is obtained by:

$$\mathcal{I}^* = \{i | p_i = 1, i = 1, \dots, N\}. \quad (14)$$

The parameter λ in Eq. (13) is used for determining whether a putative match is correct, which is also a threshold.

2.3. Guided matching strategy

There are almost no completely rigid scenes in real-world scenarios and image pairs usually undergo relatively complex non-rigid transformations due to camera distortion or noise artifacts, which will result in high ratio outliers in the putative set. In this case, neighborhood construction composed of initial putative matches becomes unreliable since the neighborhood information of feature points will be inaccurate [34].

According to the Bayesian theorem [35], the more events that support a certain attribute occur, the more likely the attribute is true. Based on this theorem, as true correspondences are consistent in local motion space, consistent matches in the local motion space are thus more likely to be true. Therefore, in this paper, we only consider a group of putative feature correspondences with low ratio outliers in the neighborhood construction. The consensus of neighborhood elements [27] are used to reject outliers from given putative correspondences. For an inlier, the distributions of its neighboring matches should be similar and quite different otherwise. Therefore, the consensus of neighborhood elements between $N_{\mathbf{x}_i}$ and $N_{\mathbf{y}_i}$ is defined as [27]:

$$r_i = \Theta_i / K \quad (15)$$

where Θ_i is the number of neighboring matches located in the two neighborhoods $N_{\mathbf{x}_i}$ and $N_{\mathbf{y}_i}$ and K is the number of nearest neighbors for the feature point under the Euclidean distance. It is easy to obtain $r_i \in [0, 1]$. Clearly, an inlier will bring a large value of r_i and vice versa. That is to say, the distributions of inliers and outliers can be distinguished by a predefined threshold α . Thus, a subset S_0 with high percentage inliers will be obtained by:

$$S_0 = \{i | r_i > \alpha\}. \quad (16)$$

In [6,10,34], the ratio test strategy of the nearest neighbor and the second nearest neighbor is used to find a subset with high ratio inliers. Nevertheless, as the number of features increases, the memory cost that stores the descriptor information of the features will be significantly enlarged. Besides, this strategy is not general enough to be suitable for every matching problem without using descriptor information (e.g. point set registration). Conversely, our proposed guided matching strategy just utilizes the spatial coordinates of the features and does not rely on the image descriptor information, therefore, it can be applied to any matching problem.

2.4. Complete ANTC method

The proposed ANTC method is summarized in Algorithm 1, which mainly contains two novel components: advanced consensus of neighborhood topology and guided matching strategy. There are five parameters in our method: K , α , λ , τ , ξ and $MaxIter$. Parameter K is the number of nearest neighbors, which consists of the neighborhood information for each feature point. Parameter α determines the precision-recall tradeoff of the subset S_0 and λ controls the threshold for finding the correctness of a correspondence. Parameter τ is used to determine whether the putative match has

high consensus of neighborhood topology with its neighboring average match. Clearly, the large value of K , τ and a small value of λ will increase the robustness simultaneously increase the time cost to some extent. In our numerical experiments, we generally choose the parameters as $K = 10$ in line 2, $K = [12, 10, 8]$ in line 7, and $\alpha = 0.5$, $\lambda = 0.8$, $\tau = 1.84$, $\xi = 0.4$, $MaxIter = 3$.

3.1. Datasets and performance measure

To comprehensively evaluate of our proposed ANTC algorithm, we verify its performance compared with other state-of-the-arts on the public datasets referring to different image types and deformations as follows:

Algorithm 1 The complete proposed ANTC method

Input: putative set $S = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$, parameters K , α , λ , τ , ξ , $MaxIter$

Output: inlier set \mathcal{I}^*

- 1 Initialize $j=0$;
- 2 Construct neighborhoods $\{N_{\mathbf{x}_i}^K, N_{\mathbf{y}_i}^K\}$ based on S .
- 3 Calculate consensus of neighborhood elements r_i using Eq.(15)
- 4 Determine a subset S_0 with high ratio inliers using Eq.(16);
- 5 $\mathcal{I}_j = S_0$
- 6 **repeat**
- 7 Construct neighborhoods $\{N_{\mathbf{x}_i}^{K_m}, N_{\mathbf{y}_i}^{K_m}\}_{i=1, m=1}^{N, M}$ based on \mathcal{I}_j ;
- 8 Calculate cost $\{\tilde{d}_i\}_{i=1}^N$ using Eq.(12);
- 9 $j=j+1$;
- 10 Determine \mathcal{I}_j using Eqs.(13) and (14);
- 11 **until** $j \geq MaxIter$;
- 12 $\mathcal{I}^* = \mathcal{I}_j$

2.5. Computational complexity

To construct K_m nearest neighbors for each feature point in S , the computational complexity of the using K-D tree is about $O((K_m + N)\log N)$. Therefore, the complexity of Lines 2 and 7 in Algorithm 1 is about $O((\sum_{m=1}^M K_m + MN)\log N)$. It is the most time-consuming step of the proposed ANTC. When the K_M neighborhood $N_{\mathbf{x}_i}^{K_M}$ is obtained, its corresponding $K_m (m < M)$ neighborhood $N_{\mathbf{x}_i}^{K_m}$ can be directly obtained from $N_{\mathbf{x}_i}^{K_M}$.

According to Eq. (12), the major cost of computing $\{\tilde{d}_i\}_{i=1}^N$ in Line 8 only involves some addition operation. Its computational complexity is less than $O(MK_M N)$. Moreover, calculating r_i , determining \mathbf{p} and \mathcal{I} using Eqs. (13) and (14) in Lines 3, 4, and 10 cost $O(N)$ complexity. There, the total time cost of the proposed ANTC is about $O(MK_M N + (K_M + N)\log N)$. The space complexity of ANTC is $O(MK_M N)$ due to the memory requirement for storing the neighborhoods $N_{\mathbf{x}_i}^{K_m}$ and $N_{\mathbf{y}_i}^{K_m}$. As $\sum_{m=1}^M K_m$ is a constant and $M \ll N$. Thus the time and space complexities of our method can be simply written as $O(N\log N)$ and $O(N)$, respectively. That is to say, our ANTC has linearithmic complexity which is significant for handling large-scale problems or real-time applications.

3. Experiments

In this section, the public available datasets and the performance measure used in this paper are introduced in Section 3.1. Then, we compare our ANTC with several state-of-the-art feature matching methods on VGG, Strecha, and remote sensing image datasets in Section 3.2. Finally, the influence of the parameters and the components are evaluated in Section 3.3.

- **VGG Datasets [36]**. VGG dataset contains 40 image pairs with five different changes in images (i.e., viewpoint changes, scale changes, image blur, JPEG compression, and illumination), which is either geometric transformation or photometric transformation. The ground truth homographies are supplied by the dataset. Similar to VFC [18], an overlap error is used to determine the match correctness.
- **Strecha Datasets [37]**. Strecha dataset contains several wide-baseline images, which is used for the multi-view stereo evaluation. We create a total of 40 image pairs for the matching evaluation involving two sceneries. These images are twenty-fifth size of the original image to accommodate slower algorithms and the first image is designated as reference. We manually check the ground truth for each putative match in each image pair to ensure objectivity.
- **Remote Sensing Image Datasets [34]**. The dataset contains 129 image pairs involving unmanned aerial vehicle images (UAV), synthetic aperture radar images, color infrared aerial photographs, and panchromatic aerial photographs. The image deformations include rigid transformation, severe noise, projective distortions, ground relief variations and large viewpoint changes. The ground truth is given by [34].

The performance measures, *Precision*, *Recall*, and *F-score*, are considered as evaluation indexes which are defined as follows

$$Precision = \frac{\text{the true positive matches number}}{\text{the preserved match number}}$$

$$Recall = \frac{\text{the true positive matches number}}{\text{the actually correct match number}}$$

$$F\text{-score} = 2 \times \frac{Precision \times Recall}{(Precision + Recall)}$$

F-score is the comprehensive measure of the performance. The matching method, which can stably find inliers from the contaminated correspondences, has a high F-score. The initial putative matches with SIFT are established by using the open source VLFeat toolbox [38]. The K-D tree is used to construct the K nearest neighbors. The experiments are performed in MATLAB R2018a on Windows 10 operating system and run on an Intel(R) 2.5-GHz machine with 4G RAM.

3.2. Experiments on feature matching

To verify the effectiveness of our ANTC with different image types, we first test it on the VGG, Strecha, and remote sensing image datasets. Our ANTC is compared with six state-of-the-arts: RANSAC [13], ICF [17], GS [22], VFC [18], LPM [32], LMR [27]. In particular, RANSAC is a traditional sampling-based method; ICF and VFC are non-parametric-based interpolation methods; GS is a graph matching method; LPM is a locality-neighborhood-based method; and LMR is a learning-based method. We implement them based on public available codes and all the parameters are fixed throughout the experiments.

Fig. 2 illustrates the initial inlier ratio with respect to VGG, Strecha and remote sensing datasets. Fig. 3 shows the quantitative comparisons of the seven methods for the VGG, Strecha, and remote sensing image datasets. Table 1 gives the average precisions, recalls, and F-scores on three datasets. For the VGG dataset, the average inlier ratio is about 32.74%, and the average number of putative match is about 2344.75. For the Strecha dataset, the average inlier ratio is about 40.32%, and the average number of putative match is about 556.63. For the remote sensing image, the average inlier ratio is about 20.78%, and the average number of putative match is about 868.85. Carefully observing Fig. 3 and Table 1, we can see that RANSAC gives very low precision. This is because the high outlier ratio increases the difficulty of obtaining an all-inlier minimal subset for transformation estimation. For the VGG dataset, ICF obtains high precision, whereas for Strecha and remote sensing dataset, it gives low precision. ICF has low recall and F-score for three datasets. GS obtains high precision but low recall and F-score. VFC achieves low precision but high recall. LPM and LMR give average precision, recall, and F-score. Our ANTC obtains the best precision-recall trade-off. This confirms the high performance of the proposed advanced consensus of neighborhood topology and the guided matching strategy.

In order to estimate the computation efficiency of different feature matching methods, the time costs (seconds) for difference methods on the three datasets are shown in Fig. 4. The average run time (seconds) of the different approaches on three datasets are listed in Table 2. From Fig. 4 and Table 2, we can see that the

fastest matching method is LPM since it takes only linearithmic time. Our proposed method is also very effective. For the VGG dataset, the average runtime of the proposed method is only about 0.0290 s while LPM is about 0.0227 s. The same results are showed by another two datasets experiments.

3.3. Influence of the components and the parameters

To verify the effectiveness of each part of our proposed ANTC algorithm, we evaluate the influence of the various values of threshold τ , the advanced consensus of neighborhood topology, and the guided matching strategy. We denote the proposed ANTC using primitive consensus of neighborhood topology defined in LPM as ANTC-P, and the proposed ANTC without using the guided matching strategy as ANTC-W.

3.3.1. Influence of various values of threshold τ

In this section, we evaluate the influence of threshold τ . The threshold τ is used to determine whether the putative match has high consensus of neighborhood topology with its neighboring average match and its value depends on three parameters (i.e. R_t , θ_t , and $\hat{\sigma}$). Usually, according to the motion coherence theory [33], in the local neighborhood of the putative match, the difference both in length and angle between \mathbf{v}_i and the average vector $\bar{\mathbf{v}}_j$ is not large for physical constraints. Therefore, in experiments, we set R_t from 0 to 1 with interval equal to 0.2 and θ_t with values $0, \frac{\pi}{12}, \frac{\pi}{6}, \frac{\pi}{4}, \frac{\pi}{3},$ and $\frac{\pi}{2}$, respectively. The aim of $\hat{\sigma}$ is used to enlarge the influence that the change of the length and angle will cause to the consensus of neighborhood topology and we empirically set $\hat{\sigma} = 0.5$. Table 3 gives the average precision-recall-F-score pairs on the VGG dataset with different values of R_t and θ_t . From Table 3, we can find that the average precision reduces with increased values of R_t or θ_t and the average recall rises. As shown in Table 3, when $R_t = 0.2$ and $\theta_t = \frac{\pi}{6}$, the highest F-score can be obtained. We substitute $R_t = 0.2$ and $\theta_t = \frac{\pi}{6}$ into Eq. (9), and then obtain the optimal value of $\tau = 1.84$, which is the value of τ used for our experiments throughout this paper.

3.3.2. Influence of advanced consensus of neighborhood topology

We present the intuitionistic matching results of ANTC-P compared with the complete ANTC in Fig. 5. The initial inlier ratio of the Image1, Image2, Image3 are 47.73%, 62.93%, 71.07%, and the match number are 1741, 1025, 999, respectively. We report the specific precision-recall-F-score statistics pairs of ANTC-P and ANTC in Table 4. From Fig. 5 and Table 4, we can find that ANTC achieves almost perfect precision-recall tradeoff. The proposed advanced consensus of neighborhood method can enlarge the distributions of outliers and inliers making it easy to identify outliers.

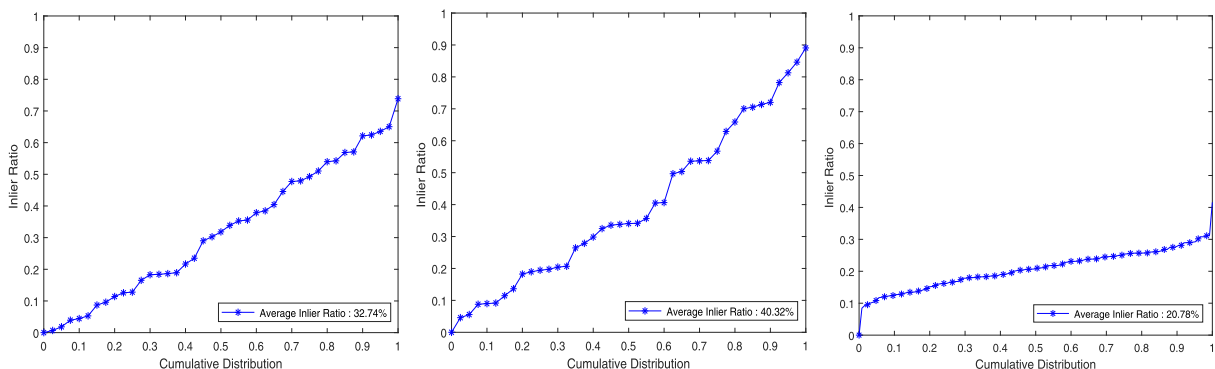


Fig. 2. Average inlier ratio. From left to right column: the initial inlier ratio of the three datasets VGG, Strecha and Remote sensing Datasets regarding the cumulative distribution. A point with coordinates (x, y) on a curve indicates there are (100*x)% of image pairs, and their inlier ratio does not exceed y.

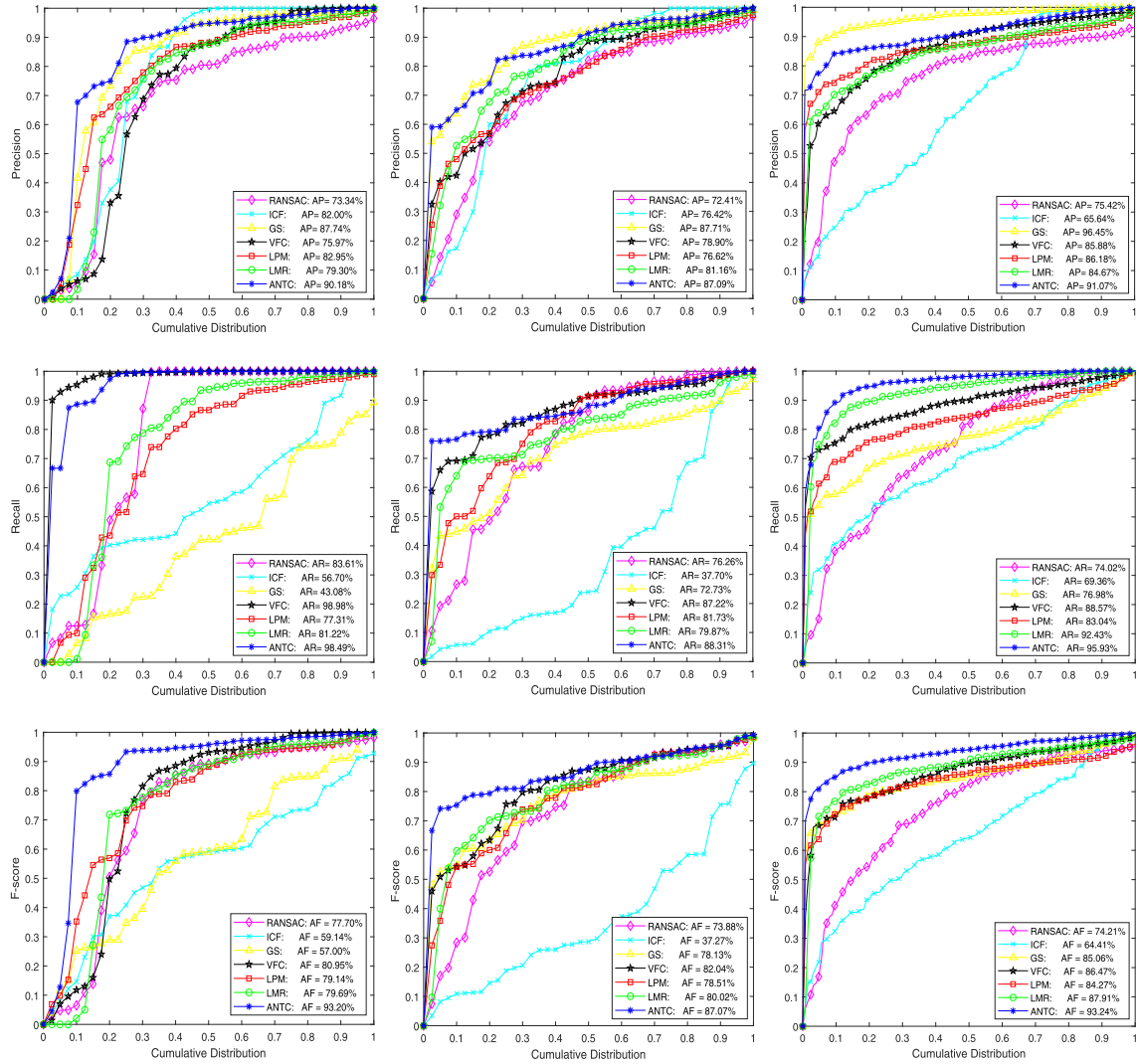


Fig. 3. Quantitative comparisons of RANSAC [13], ICF [17], GS [22], VFC [18], LPM [32], LMR [27], and our ANTC. From left to right column: VGG, Strecha and Remote sensing Datasets. From the top row to the bottom row: precision, recall, and F-score with regard to the cumulative distribution.

Table 1

Average precision-recall-F-scores pairs (%) of different matching methods for three datasets.

Method	VGG	Strecha	RS
RANSAC [13]	(73.05, 82.03, 76.68)	(72.41, 76.26, 73.88)	(75.42, 74.02, 74.21)
ICF [17]	(82.00, 56.70, 59.14)	(76.42, 37.70, 37.27)	(65.64, 69.36, 64.41)
GS [22]	(87.74, 43.08, 57.00)	(87.71, 72.73, 78.13)	(96.45, 76.98, 85.06)
VFC [18]	(72.22, 98.69 , 77.99)	(78.90, 87.22, 82.04)	(85.88, 88.57, 86.47)
LPM [32]	(82.95, 77.31, 79.14)	(76.62, 81.73, 78.51)	(86.18, 83.04, 84.27)
LMR [27]	(79.03, 81.22, 79.69)	(81.16, 79.87, 80.02)	(84.67, 92.43, 87.91)
Ours	(90.18, 98.49, 93.20)	(87.09, 88.31, 87.07)	(91.07, 95.93, 93.24)

3.3.3. Influence of guided matching strategy

We also present the intuitionistic matching results of ANTC-W compared with the complete ANTC in Fig. 6. To make the test more challenging, we choose another three image pairs with low inlier ratio. When the image pairs undergo complex transformation (e.g. large viewpoint changes in *Image4*, wide baseline in *Image5*, and small overlap in *Image6*), the whole putative set will contain large outlier proportion. The initial inlier ratio of the *Image4*, *Image5*, *Image6* are 18.28%, 27.87%, 19.56%, and the match number are 2478, 689, and 869, respectively. In this case, the neighborhood

construction based on the whole putative set is unreliable. Table 5 gives the specific precision-recall-F-score statistics pairs of ANTC-W and ANTC. From Fig. 6 and Table 5, we can find that our guided matching strategy can effectively improve the matching results, especially for recall- and F- score.

3.4. Robustness analysis

In order to investigate the generality of our proposed method, we collect 38 image pairs referring to different features and objects

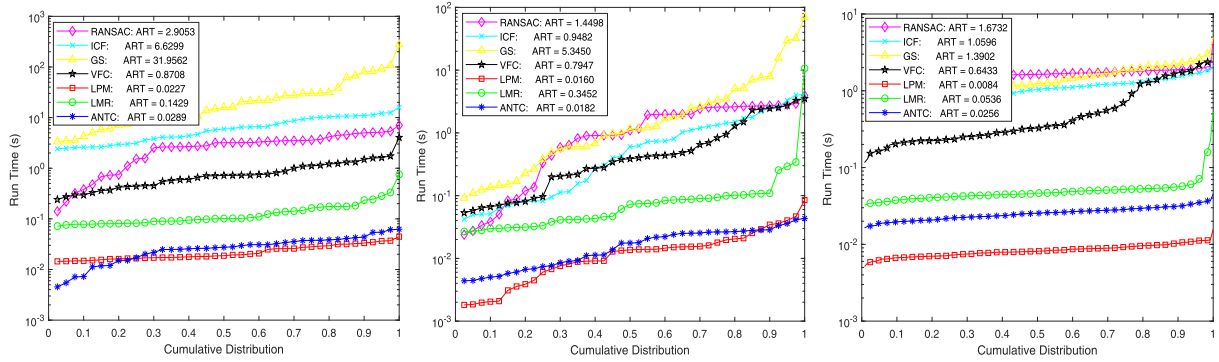


Fig. 4. Comparison of computation efficiency of different methods on the three datasets. (Left to right) VGG, Strecha and Remote sensing Datasets.

Table 2

Average run time (ART) comparison on three datasets.

	RANSAC [13]	ICF [17]	GS [22]	VFC [18]	LPM [32]	LMR [27]	Ours
VGG	2.9053	6.6299	31.956	0.8708	0.0227	0.1492	0.0290
Strecha	1.4498	0.9482	5.3450	0.7947	0.0160	0.3452	0.0182
RS	1.6732	1.0596	1.3902	0.6433	0.0084	0.0536	0.0256

Table 3

Average precisions, recalls and F-scores (%) on the VGG dataset with respect to different values of θ_t and R_t .

$\theta_t \setminus R_t$	0	0.2	0.4	0.6	0.8	1
0	(99.87 , 76.85, 86.86)	(97.32, 84.23, 90.30)	(94.25, 87.32, 90.65)	(92.76, 92.43, 92.59)	(88.18, 98.49, 93.05)	(80.24, 99.02, 88.65)
$\frac{\pi}{12}$	(97.32, 84.23, 90.30)	(94.25, 87.32, 90.65)	(91.78, 92.43, 92.10)	(86.54, 99.02, 92.36)	(80.24, 99.02, 88.65)	(72.95, 99.12, 84.04)
$\frac{\pi}{6}$	(93.97, 88.32, 91.06)	(90.18, 98.49, 94.15)	(85.28, 99.02, 91.64)	(79.47, 99.12, 88.21)	(70.23, 99.26, 82.26)	(69.31, 99.32, 81.64)
$\frac{\pi}{4}$	(88.18, 98.49, 93.05)	(83.56, 99.02, 90.64)	(75.44, 99.12, 85.67)	(70.23, 99.26, 82.26)	(67.78, 99.56 , 80.65)	(63.25, 99.56 , 77.36)
$\frac{\pi}{3}$	(80.24, 99.02, 88.65)	(72.95, 99.12, 84.04)	(69.31, 99.32, 81.64)	(67.21, 99.56 , 80.25)	(62.88, 99.56 , 77.08)	(60.93, 99.56 , 75.60)
$\frac{\pi}{2}$	(67.78, 99.56 , 80.65)	(63.25, 99.56 , 77.36)	(60.93, 99.56 , 75.60)	(56.32, 99.56 , 71.94)	(56.32, 99.56 , 71.94)	(56.32, 99.56 , 71.94)

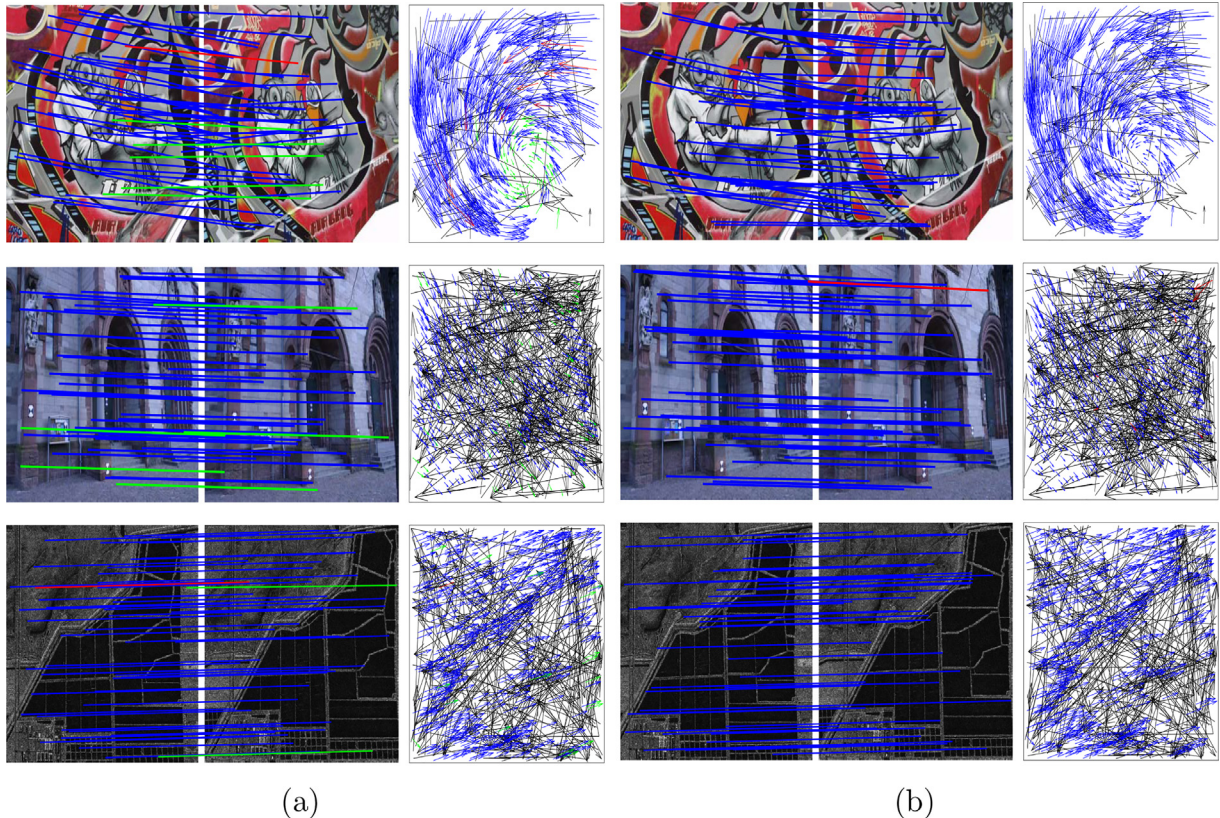


Fig. 5. Qualitative results of ANTC-P and ANTC on three typical image pairs from the three datasets: (Top to bottom) VGG (*Image1*), Strecha (*Image2*), and RS (*Image3*). (a) Mismatch removal result of ANTC-P. (b) Mismatch removal result of ANTC. The lines in the image pairs are the matching results with respect to the arrows in the motion field (red = false positive, green = false negative, blue = true positive and black = true negative). For visibility, we randomly selected no more than 50 matches to be presented, and do not show the true negative.

Table 4
Precision-recall-F-score pairs (%) of ANTC-P and ANTC on three image pairs.

	Image1	Image2	Image3
ANTC-P	(98.61, 91.23, 94.77)	(100.0 , 91.01, 95.29)	(99.86 , 96.76, 98.28)
ANTC	(100.0 , 100.0 , 100.0)	(98.77, 100.0 , 99.38)	(98.88, 99.86 , 99.37)

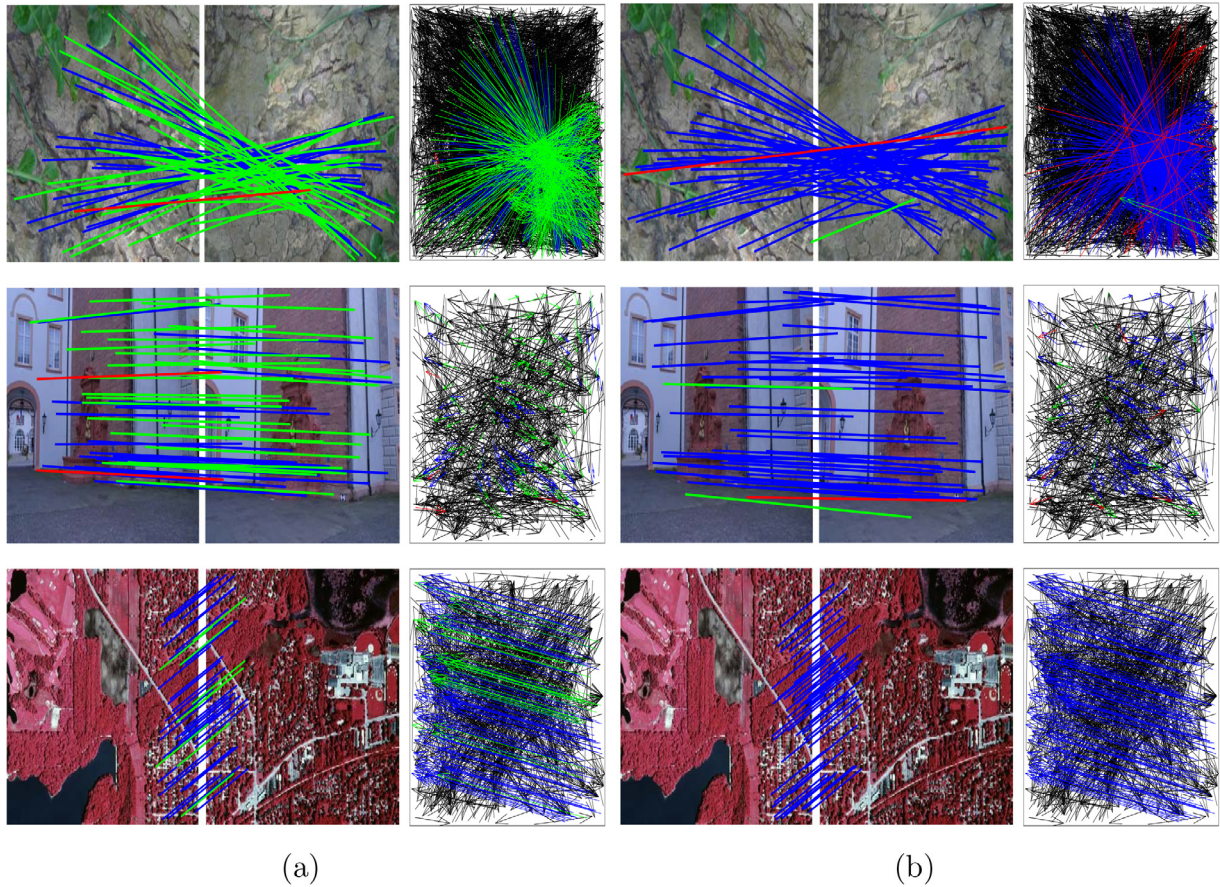


Table 5
Precision-recall-F-score pairs (%) of ANTC-W and ANTC on three image pairs.

	Image4	Image5	Image6
ANTC-W	(98.38 , 40.18, 57.05)	(96.00, 50.00, 65.75)	(100.0 , 80.59, 89.25)
ANTC	(94.75, 99.56 , 97.09)	(96.05 , 88.54 , 92.14)	(100.0 , 100.0 , 100.0)

including homes [39], architectures [40], vehicles and nature [36]. We show the examples of the image pairs above in Fig. 7. The average inlier ratio of homes, architectures, vehicles and nature are 42.3%, 35.7%, 27.4% and 45.4%, respectively. We test the performance of our method and present the average precision, recall and F-score with respect to these object types in Table 6. From Table 6, we can find that our proposed method achieves satisfactory matching results since the basic formulation is general enough to deal with different object matching problems. Actually, the performance of outlier elimination techniques depends slightly on the

features and objects and it may be affected by the correspondence distribution in the image pair.

4. Conclusion

In this paper, we propose a simple yet efficient feature matching method based on advanced neighborhood topology consensus termed as ANTC. We formulate the feature matching problem into a mathematical model and adopt an advanced consensus of



Fig. 7. Examples of the image pairs with respect to different objects.

Table 6

Precision-recall-F-score pairs (%) of the proposed method on different object types.

	Homes	Architectures	Vehicles	Nature
ANTC	(97.44, 96.88, 97.37)	(98.89, 99.36, 98.99)	(97.34, 99.38, 98.75)	(98.47, 97.05, 98.21)

neighborhood topology to further exploit the topological structure of the matches. Specifically, the use of the advanced consensus of neighborhood topology can robustly recover the inliers from the contaminated correspondences. An efficient guided matching strategy can help to filter out most of the outliers and obtain a subset with high ratio inliers, which can significantly boost the true matches without sacrifice in accuracy. The experimental results on various image types for feature matching demonstrate that the proposed method outperforms the state-of-the-art methods.

In future work, to better introduce the proposed method to high dimensional data analysis [41–43] (e.g. three-dimensional point cloud registration, medical images registration) or Internet of Things networks [44,45], both motion consistency and advanced neighborhood topology consensus should be better defined. Moreover, the proposed method can be integrated into other well-known feature matching pipelines to investigate the better combinations. The code of our method is at: <https://github.com/lyz8023lyp/ANTC>.

CRedit authorship contribution statement

Yizhang Liu: Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing. **Yanping Li:** Methodology, Software. **Luanyuan Dai:** Software, Visualization. **Changcai Yang:** Supervision, Conceptualization, Methodology, Writing - original draft, Writing - review & editing. **Lifang Wei:** Writing - review & editing. **Taotao Lai:** Writing - review & editing. **Riqing Chen:** Writing - review & editing, Funding acquisition.

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.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 61701117, 61702101, 61802064, and 61501120, in part by the Natural Science Fund of Fujian Province under Grant 2019J01402, in part by the Fujian Province Health Education Joint Research Project under Grant 2019WJ28, and in part by the Science and Technology Innovation Special Fund Project of Fujian Agriculture and Forestry University under Grant CXZX2019124S and CXZX2019125G.

References

- [1] L. Liang, W. Zhao, X. Hao, Y. Yang, K. Yang, L. Liang, Q. Yang, Image registration using two-layer cascade reciprocal pipeline and context-aware dissimilarity measure, *Neurocomputing* 371 (2020) 1–14.
- [2] Robust and efficient gmm-based free-form parts registration via bi-directional distance, *Neurocomputing* 360 (2019) 279–293.
- [3] S. Tang, W. Cong, J. Yang, T. Fu, H. Song, D. Ai, Y. Wang, Local statistical deformation models for deformable image registration, *Neurocomputing* 303 (2018) 1–10.
- [4] J. Ma, C. Chen, C. Li, J. Huang, Infrared and visible image fusion via gradient transfer and total variation minimization, *Information Fusion* 31 (2016) 100–109.
- [5] J. Ma, Y. Ma, C. Li, Infrared and visible image fusion methods and applications: A survey, *Information Fusion* 45 (2019) 153–178.
- [6] W.-Y. Lin, F. Wang, M.-M. Cheng, S.-K. Yeung, P.H. Torr, M.N. Do, J. Lu, Code: Coherence based decision boundaries for feature correspondence, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (1) (2017) 34–47.
- [7] I. Carlbom, D. Terzopoulos, K.M. Harris, Computer-assisted registration, segmentation, and 3d reconstruction from images of neuronal tissue sections, *IEEE Transactions on Medical Imaging* 13 (2) (1994) 351–362.
- [8] C. Yang, M. Zhang, Z. Zhang, L. Wei, R. Chen, H. Zhou, Non-rigid point set registration via global and local constraints, *Multimedia Tools and Applications* 77 (24) (2018) 31607–31625.
- [9] C. Yang, Y. Liu, X. Jiang, Z. Zhang, L. Wei, T. Lai, R. Chen, Non-rigid point set registration via adaptive weighted objective function, *IEEE Access* 6 (2018) 75947–75960.
- [10] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* 60 (2) (2004) 91–110.

- [11] J. Ma, X. Jiang, A. Fan, J. Jiang, J. Yan, Image matching from handcrafted to deep features: A survey, *International Journal of Computer Vision* (2020) 1–57, <https://doi.org/10.1007/s11263-020-01359-2>.
- [12] X. Jiang, J. Ma, J. Jiang, X. Guo, Robust feature matching using spatial clustering with heavy outliers, *IEEE Transactions on Image Processing* 29 (2020) 736–746.
- [13] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM* 24 (6) (1981) 381–395.
- [14] G. Xiao, H. Wang, Y. Yan, D. Suter, Superpixel-guided two-view deterministic geometric model fitting, *International Journal of Computer Vision* 127 (4) (2019) 323–339.
- [15] E. Brachmann, A. Krull, S. Nowozin, J. Shotton, F. Michel, S. Gumhold, C. Rother, Dsac-differentiable ransac for camera localization, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6684–6692.
- [16] D. Barath, J. Matas, Graph-cut ransac, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6733–6741.
- [17] X. Li, Z. Hu, Rejecting mismatches by correspondence function, *International Journal of Computer Vision* 89 (1) (2010) 1–17.
- [18] J. Ma, J. Zhao, J. Tian, A.L. Yuille, Z. Tu, Robust point matching via vector field consensus, *IEEE Transactions on Image Processing* 23 (4) (2014) 1706–1721.
- [19] J. Ma, W. Qiu, J. Zhao, Y. Ma, A.L. Yuille, Z. Tu, Robust L2E estimation of transformation for non-rigid registration, *IEEE Transactions on Signal Processing* 63 (5) (2015) 1115–1129.
- [20] W.-Y.D. Lin, M.-M. Cheng, J. Lu, H. Yang, M. N. Do, P. Torr, Bilateral functions for global motion modeling, in: *European Conference on Computer Vision*, Springer, 2014, pp. 341–356.
- [21] J. Ma, J. Wu, J. Zhao, J. Jiang, H. Zhou, Q.Z. Sheng, Nonrigid point set registration with robust transformation learning under manifold regularization, *IEEE Transactions on Neural Networks and Learning Systems* 30 (12) (2019) 3584–3597.
- [22] H. Liu, S. Yan, Common visual pattern discovery via spatially coherent correspondences, in: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 2010, pp. 1609–1616.
- [23] K. Adamczewski, Y. Suh, K. Mu Lee, Discrete tabu search for graph matching, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 109–117.
- [24] P. Swoboda, A. Mokarian, C. Theobalt, F. Bernard, et al., A convex relaxation for multi-graph matching, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11156–11165.
- [25] C. Wang, L. Wang, L. Liu, Progressive mode-seeking on graphs for sparse feature matching, in: *European Conference on Computer Vision*, Springer, 2014, pp. 788–802.
- [26] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [27] J. Ma, X. Jiang, J. Jiang, J. Zhao, X. Guo, Lmr: Learning a two-class classifier for mismatch removal, *IEEE Transactions on Image Processing* 28 (8) (2019) 4045–4059.
- [28] K. Moo Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, P. Fua, Learning to find good correspondences, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2666–2674.
- [29] C. Zhao, Z. Cao, C. Li, X. Li, J. Yang, Nm-net: Mining reliable neighbors for robust feature correspondences, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 215–224.
- [30] Cross-weather image alignment via latent generative model with intensity consistency, *IEEE Transactions on Image Processing* 29 (2020) 5216–5228.
- [31] J. Bian, W.-Y. Lin, Y. Matsushita, S.-K. Yeung, T.-D. Nguyen, M.-M. Cheng, Gms: grid-based motion statistics for fast, ultra-robust feature correspondence, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4181–4190.
- [32] J. Ma, J. Zhao, J. Jiang, H. Zhou, X. Guo, Locality preserving matching, *International Journal of Computer Vision* 127 (5) (2019) 512–531.
- [33] A.L. Yuille, N.M. Grzywacz, A mathematical analysis of the motion coherence theory, *International Journal of Computer Vision* 3 (2) (1989) 155–175.
- [34] J. Ma, J. Jiang, H. Zhou, J. Zhao, X. Guo, Guided locality preserving feature matching for remote sensing image registration, *IEEE Transactions on Geoscience and Remote Sensing* 56 (8) (2018) 4435–4447.
- [35] J. Joyce, Bayes' Theorem, in: E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2019 Edition, Metaphysics Research Lab, Stanford University, 2019.
- [36] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, A comparison of affine region detectors, *International Journal of Computer Vision* 65 (1) (2005) 43–72.
- [37] C. Strecha, W. Von Hansen, L. Van Gool, P. Fua, U. Thoennessen, On benchmarking camera calibration and multi-view stereo for high resolution imagery, in: *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2008, pp. 1–8.
- [38] L. Torresani, V. Kolmogorov, C. Rother, Feature correspondence via graph matching: Models and global optimization, in: *European Conference on Computer Vision*, Springer, 2008, pp. 596–609.
- [39] H. Aanæs, R.R. Jensen, G. Vogiatzis, E. Tola, A.B. Dahl, Large-scale data for multiple-view stereopsis, *International Journal of Computer Vision* 120 (2) (2016) 153–168.
- [40] E. Tola, V. Lepetit, P. Fua, Daisy: An efficient dense descriptor applied to wide-baseline stereo, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (5) (2009) 815–830.
- [41] M. Woźniak, D. Połap, Soft trees with neural components as image-processing technique for archeological excavations, *Personal and Ubiquitous Computing* (2020) 1–13.
- [42] M. Woźniak, D. Połap, Adaptive neuro-heuristic hybrid model for fruit peel defects detection, *Neural Networks* 98 (2018) 16–33.
- [43] M. Woźniak, D. Połap, Bio-inspired methods modeled for respiratory disease detection from medical images, *Swarm and Evolutionary Computation* 41 (2018) 69–96.
- [44] S.P. RM, S. Bhattacharya, P.K.R. Maddikunta, S.R.K. Somayaji, K. Lakshmana, R. Kaluri, A. Hussien, T.R. Gadekallu, Load balancing of energy cloud using wind driven and firefly algorithms in internet of everything, *Journal of Parallel and Distributed Computing*.
- [45] G.T. Reddy, M.P.K. Reddy, K. Lakshmana, R. Kaluri, D.S. Rajput, G. Srivastava, T. Baker, Analysis of dimensionality reduction techniques on big data, *IEEE Access* 8 (2020) 54776–54788.



Yizhang Liu received the B.S. degree in electronic and information engineering and master's degree in computer science and technology from Fujian Agriculture and Forestry University, Fuzhou, China, in 2013 and 2017, respectively. He is currently pursuing the Eng.D. degree with the school of software engineering, Tongji University, Shanghai. His current research interests include computer vision and image matching.



Yanping Li received the B.S. degree in computer science and technology from Fujian Agriculture and Forestry University, Fuzhou, China, in 2017, and her master's degree in the computer science and technology at Hohai University, Nanjing, China, in 2020. She is currently pursuing the Eng.D. degree with the department of computer science and technology, Tongji University, Shanghai. Her current research interests include computer vision and image matching.



Luanyuan Dai received the B.S. degree in Geophysics from Northeast Petroleum University, Daqing, China in 2014. She is currently studying the master's degree with the School of computer and information, Fujian Agriculture and Forestry University, Fuzhou, China. Her current research interests mainly include computer vision and image matching.



Changcai Yang received M.S. degree in control theory and control engineering from China Three Gorges University, China, in 2008, and his Ph.D. in pattern recognition and intelligent systems at Huazhong University of Science and Technology (HUST), China, in 2012. From 2012 to 2014, he worked as Post Doctor at HUST, China. Since 2016, he has been an Associate Professor and MS Supervisor with the College of Computer and Information Science, Fujian Agriculture and Forestry University, Fuzhou, China. He has authored or coauthored more than 40 papers. His research interests include computer vision, image processing, point set registration.



Lifang Wei is an associate Professor with College of Computer and Information, Fujian Agriculture and Forestry University, Fuzhou, China. She received the bachelor degree in Electronic Communication engineering from Xi'an Shiyou University on 2005, Xi'an, China and received the master degree in biomedical engineering from Northwest Polytechnical University on 2008, Xi'an, China. She had received her Ph.D in Communication and Information System from Fuzhou University on 2013, Fuzhou, China. Her research interests focus on computer vision and image processing.



Riqing Chen received the B.Eng. degree of communication engineering from Tongji University, China, in 2001, the M.Sc. degree of communications and signal processing from Imperial College London, U.K., in 2004, and the D.Phil. degree of engineering science from the University of Oxford, U.K., in 2010. Since 2014, he has been affiliated with Digital Fujian Institute of the Big Data for Agriculture and Forestry, Fujian Agriculture and Forestry University, Fuzhou, China. His research interests include computer vision, big data and visualization, cloud computing, consumer electronics, flash memory, and wireless sensor networking etc.



Taotao Lai received the Ph.D. degree in computer science and technology from Xiamen University, Xiamen, China, in 2016. From 2016 to 2018, he worked as Post Doctor at Fujian Agriculture and Forestry University, China. He is currently a lecturer in College of Computer and Control Engineering, Minjiang University, Fuzhou, China. He has authored or coauthored several papers in international journals, including the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, Pattern Recognition, Computer Vision and Image Understanding, etc. His research interests include robust model fitting and structure from motion.