Image matching via the local neighborhood for low inlier ratio

Weiqing Wang,[a,b] **Yongrong Sun**,[a,*] **Zhong Liu**,[b] **Zhantian Qin**,[b] **Can Wang**,[b] **and Jinchang Qin**[b]

[a]Nanjing University of Aeronautics and Astronautics, Navigation Research Center, College of Automation Engineering, Nanjing, China

[b]Guilin University of Aerospace Technology, School of Mechanical and Electrical Engineering, Guilin, China

Abstract. Establishing reliable correspondences in an image pair is a prerequisite and crucial in computer vision. It remains a difficult topic to separate true and false matches in the given putative dataset with a low inlier ratio. To address this problem, an image matching method is proposed, via the local neighborhood of feature points. Grid-based motion statistics are initially engaged to preprocess the putative dataset, especially in which the inlier ratio is low. The local neighborhood distributions of feature points are then collected as the quality function of correspondences. Progressive sample consensus is next employed to estimate a global deformation for removing false matches. Robust experiments on nine typical image pairs with low inlier ratios demonstrate the superiority of our proposed method over five state-of-the-art methods. The comparison experiments on the Oxford dataset illustrate that our method outperforms the other five image matching methods. [DOI: 10.1117/1.JEI.31.2.023039]

Keywords: image matching; mismatch removal; local neighborhood; grid-based motion statistics; progressive sample consensus.

Paper 210886G received Nov. 30, 2021; accepted for publication Apr. 6, 2022; published online Apr. 29, 2022.

1 Introduction

Establishing reliable correspondences in an image pair is a fundamental problem in computer vision, and it has been a crucial prerequisite for many computer vision tasks, such as image restoration,[1] image retrieval,[2] image registration,[3,4] point set registration,[5] and image search.[6] For example, the detailed quality of a multispectral and multimodal image registration depends on the performance of matching.

Image matching typically consists of a two-stage strategy.[7,8] A putative dataset is initially generated, and false matches are then removed from them. In general, the putative dataset is constructed by some feature extraction with their similar feature descriptors [e.g., scale-invariant feature transform (SIFT)[9] and oriented fast and rotated BRIEF[10] (ORB)][11]. As the local descriptor is ambiguous, the putative dataset is contaminated by a large percentage of false matches. When the image pairs have significant viewpoint changes, heavy occlusion, scale, and rotation changes, the matching difficulty becomes considerably worse. Hence, a designed robust method plays an important role in establishing reliable correspondences.

Due to simplicity and robustness, random sample consensus (RANSAC)[12] is most popularly engaged to estimate a parametric model for removing mismatches. RANSAC is a hypothesize-and-verify method, attaining to find the smallest possible outlier-free subset to estimate a given parametric model by resampling randomly. During the past few decades, some varieties of RANSAC have been developed to address the false matches removal problem. Nevertheless, it remains a challenge to attain an effective and efficient method for practical use. The challenges mainly come from three aspects. First, a predefined parametric model is needed to construct the

*Address all correspondence to Yongrong Sun, sunyr@nuaa.edu.cn

relationship of correspondences. When the underlying image transformation is nonrigid, it becomes less efficient. Second, it plays an important role in choosing an appropriate threshold. Based on the threshold, it separates true matches and false matches. The threshold strongly influences the outcome of RANSAC. In general, a threshold is given through experience. Third, a high attraction is decreasing the time consuming of attaining a good deformation. Its running time is related to the inlier ratio, the evidence, and the minimum parameters required for model estimation. When the inlier ratio becomes low, it has a long-running time and tends to severely degrade.

To remove false matches in the putative dataset, several nonparametric methods are proposed. Through presuming the motion smoothness constraints, grid-based motion statistics (GMS)[13] quickly eliminates outliers with statistical measures based on the number of neighboring matches. Locality preserving matching (LPM)[14] maintains the spatial neighborhood association among feature points reflecting the topological structures of an image scene is generally well preserved due to physical constraints. Li et al.[15] provided a new local barycentric coordinate system for feature matching and converted it into a local structure-preserving mathematical model. Ma et al.[16] transformed the matching problem into spatial clustering with outliers and engages the classic density-based spatial clustering of applications with noise to solve it. However, when the inlier ratio of the putative dataset is low, their performances of removing false matches would become worse.

To remove false matches in the given putative dataset with a low inlier ratio, an image matching method is proposed, which is via the local neighborhood of feature points. GMS are initially employed to preprocess the putative dataset for increasing its inlier ratio, especially in which the inlier ratio is lower than 25%. The local neighborhood distributions of feature points are then assembled as the quality function of correspondences. Progressive sample consensus is next adopted to estimate a global deformation for removing false matches.

More concretely, there is one main contribution to our paper. The local neighborhood distribution of feature points is added as the weight of correspondence for progressive sample consensus in image matching. Due to the motion slow and smoothness constraints, many matches are likely to be inliers in a small neighborhood of a true match.

The remainder of this paper is organized as follows. Section 2 briefly reviews past image matching methods. In Sec. 3, an image matching method via the local neighborhood is provided. Section 4 presents the experimental results and discussions. Finally, Sec. 5 provides a summary and concluding remarks.

2 Related Works

Existing incorrect matches removal works can be roughly classified into three major categories, i.e., RANSAC-like methods, nonparametric methods, and learning-based methods. We go over a few representative methods and present them below.

2.1 *RANSAC-Like Methods*

RANSAC is most popularly engaged to estimate a parametric model for removing incorrect matches. During the past few decades, numerous varieties of RANSAC have been proposed to attain better performance, such as spatial consistency on RANSAC (SC-RANSAC),[17] progressive sample consensus (PROSAC),[18] EVSAC,[19] locally optimized RANSAC (LO-RANSAC),[20] graph cut RANSAC (GC-RANSAC),[21] and MAGSAC.[22] SC-RANSAC exploits spatial relations between feature points in two images for increasing the inlier ratio. The hypothesis is generated by utilizing a few high-scored corresponded points. Similarly, Tang et al.[23] adopted GMS to filter the putative dataset, and wang et al.[24] integrated locality preserving constraints to remove mismatches by preprocessing the dataset. PROSAC samples are taken from increasingly larger sets of top-ranked correspondences, with the linear ordering determined by a similarity function on the set of correspondences. To speed up accurate hypothesis generation, EVSAC uses a probabilistic framework to assign a confidence value to each match. LO-RANSAC engaged a local optimization process to estimate the deformation.

When a current best model is found, GC-RANSAC implements the graph cut strategy by exploiting the spatial coherence in the local optimization step. MAGSAC introduces δ-consensus to eliminate the need for predefining a threshold. Furthermore, when dealing with image pairs that undergo extensive nonrigid transformations, RANSAC-like methods become less efficient.

2.2 *Nonparametric Methods*

To address the aforementioned issue, several nonparametric methods have been investigated, including vector field consensus (VFC),[25] LPM,[14] robust feature matching based on spatial clustering algorithm with noisy samples (RFM-SCAN),[16] and GMS. VFC employs maximum likelihood estimation for parameters in the mixed probabilistic model to estimate the probability of inliers, assuming that noise around inliers and outliers emanated from independent distributions. Due to physical restrictions, LPM preserves the spatial neighborhood association among feature points reflecting the topological structures of an image scene. To address the issue, RFM-SCAN transforms it into spatial clustering with outliers and adopts the classic density-based spatial clustering of applications with noise. GMS swiftly eliminates outliers using statistical measures based on the number of neighboring matches by assuming motion smoothness constraints. Hence, GMS can be adopted to preprocess the putative dataset for boosting the inlier ratio.

2.3 *Learning-Based Methods*

Learning-based methods are another technique to deal with the feature matching problem, including learning to find good correspondences (LFGC),[26] mining reliable neighbors for robust feature correspondences (NM-NET),[27] learning a two-class classifier for mismatch removal (LMR),[28] and learning feature matching graph neural networks (SuperGlue).[29] LFGC engaged a deep network to find geometrically consistent correspondences. For correspondence selection, NM-Net presented a deep classification network that fully mined compatibility-specific locality. LMR embraces a supervised learning technique to learn a two-class classifier through the consensus of local neighborhood structures. In SuperGlue, finding the partial assignment between two sets of local features is considered as learning feature matching. Even though deep learning architectures generate better representations than handcrafted representations, the putative set still contains a significant number of mismatches.

3 Method

Before introducing our proposed method, the gird-based motion statistics are initially employed to increase the inlier ratio proportion of the putative dataset. The progressive sample consensus is then engaged to boost the process of estimating deformation. The feature point distribution in the local neighborhood of a certain feature point is investigated in the introduction of our proposed method, and it plays a significant role in image matching.

3.1 *Grid-Based Motion Statistics*

A GMS algorithm is a nonparametric method of image matching, which rapidly removes the mismatches from the putative dataset. GMS assumes that the motion is slow and smooth. It presumes that many correspondences are more likely to be true in a small neighborhood of correct matching points due to motion smooth. Conversely, abundant correspondences are more likely to be the outliers which are in the neighborhoods of incorrect matching points.

Suppose that image pair $\{A, B\}$ have $\{M, Q\}$ feature points, respectively. $C = \{c_1, c_2, \ldots, c_i, \ldots, c_M\}$ is the set of all nearest-neighborhood feature correspondences from A to B. C has cardinality $|C| = M$. It analyzes the local support of each correspondence to divide C into sets of true matches and mismatches. Given $\{a, b\}$ is one region in the image pair $\{A, B\}$, which have $\{m, q\}$ feature points, respectively. d_a indicates one of the m feature points in the region a. Assume the matching probability of each feature is independent and given d_a has probability t of matching correctly. When the matching is false, its nearest-neighbor match can lie in any of the Q possible locations. Thus,

$$P(d_a^b|d_a^f) = fq/Q, \tag{1}$$

where d_a^b is one correspondence from the region a to the region b, and d_a^f denotes that the match is false. f is a factor.

Given T^{ab} indicates that the regions $\{a, b\}$ view the same location. Let $P_t = P(d_a^b|T^{ab})$ be the probability that, given $\{a, b\}$ view the same location, feature d_a's nearest neighbor is in region b. Thus,

$$\begin{aligned} P_t &= P(d_a^b|T^{ab}) \\ &= P(d_a^t|T^{ab}) + P(d_a^f, d_a^b|T^{ab}), \end{aligned} \tag{2}$$

where d_a^t indicates that the match is true. As the matching probability of each feature is independent, there is the following equation:

$$P(d_a^t|T^{ab}) = P(d_a^t) = t. \tag{3}$$

Through Bayesian rule,

$$\begin{aligned} P(d_a^f, d_a^b|T^{ab}) &= P(d_a^f|T^{ab})P(d_a^b|d_a^f, T^{ab}) \\ &= P(d_a^f)P(d_a^b|d_a^f) \\ &= \frac{(1-t)fq}{Q}. \end{aligned} \tag{4}$$

By Eqs. (3) and (4), Eq. (2) can be obtained

$$\begin{aligned} P_t &= P(d_a^b|T^{ab}) \\ &= t + \frac{(1-t)fq}{Q}. \end{aligned} \tag{5}$$

Given F_a^b denotes that the regions $\{a, b\}$ view the different location. Let $P_f = P(d_a^b|F^{ab})$ be the probability that, given $\{a, b\}$ view the different locations, feature d_a's nearest neighbor is in region b. The similar processing method as Eq. (2), P_f can be obtained,

$$\begin{aligned} P_f &= P(d_a^b|F^{ab}) \\ &= P(d_a^f, d_a^b|F^{ab}) \\ &= P(d_a^f|F^{ab})P(d_a^b|d_a^f, F^{ab}) \\ &= P(d_a^f)P(d_a^b|d_a^f) \\ &= \frac{(1-t)fq}{Q}. \end{aligned} \tag{6}$$

Given S_i is a measure of neighborhood support:

$$S_i = |C_i| - 1, \tag{7}$$

where $C_i \subseteq C$ is the subset of matches between regions $\{a, b\}$ of match c_i. Since the matching of each feature is independent, using Eqs. (5) and (6), it can be approximated that the distribution of S_i, the number of matches in a neighborhood of match c_i, with a pair of binomial distribution:

$$S_i \sim \begin{cases} B(n, P_t), & \text{if } c_i \text{ is true} \\ B(n, P_f), & \text{if } c_i \text{ is false} \end{cases}, \tag{8}$$

where n is the number of matches in a neighborhood of match c_i. Through Eq. (8), compared with the true matches, the distribution of the neighborhood scores S of the false matches has a different distribution. The probability distributions of neighborhood scores of true and false

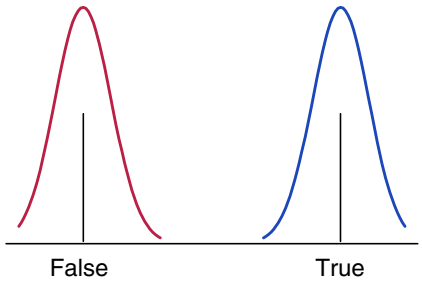

Fig. 1 Probability distributions of neighborhood scores of true and false matches.

matches are shown in Fig. 1. Therefore, a threshold is chosen to separate true matches from the putative dataset.

In GMS, the images A and B are divided into 20×20 nonoverlapping grids. For each grid (cell) in A or B, the grid contains the maximum amount of correspondences. S_{ab} is a measure of neighborhood support in cell-pair $\{a, b\}$. S_{ab} can be estimated as

$$S_{ab} = \sum_{k=1}^{k=9} |C_{a^k b^k}|, \tag{9}$$

where $|C_{a^k b^k}|$ is the amount of correspondences in the cell-pair $\{a^k, b^k\}$. All correspondences in cell-pair $\{a, b\}$ are considered as inliers if

$$\text{cell} - \text{pair}\{a, b\} \in \begin{cases} \text{True}, & \text{if } S_{ab} > \tau_a = \alpha\sqrt{n_a} \\ \text{False}, & \text{otherwise} \end{cases}, \tag{10}$$

where the true indicates that the correspondences are inliers. Conversely, the false represents that the correspondences are outliers. τ_a is the threshold approximated by $\alpha\sqrt{n_a}$. The α is a given parameter and n_a is the average amount of correspondences.

3.2 *Progressive Sample Consensus*

Progressive sample consensus is a variant of RANSAC that regards the impact of sample probability. The dataset is initially sorted tentative matches in descending order by a quality function e:

$$c_i, c_j \in C : i < j \Rightarrow e(c_i) \geq e(c_j). \tag{11}$$

A sample m is a subset of M tentative matches, and its quality function is defined as the lowest quality of a match included in the sample:

$$e(m) = \min_{c_i \in C_m} e(c_i). \tag{12}$$

After the samples of size m out of M data points are fulfilled, the samples are sorted in descending order according to Eq. (12).

Let T_q be an average number of samples from the tentative dataset which is a set of q matches:

$$T_q = T_M \frac{\binom{q}{m}}{\binom{M}{m}} = T_M \prod_{i=0}^{m-1} \frac{q-i}{M-i}, \tag{13}$$

where T_M is an average number of samples from the entire dataset. By Eq. (13), T_{q+1} could be recursively defined as

$$T_{q+1} = T_M \prod_{i=0}^{m-1} \frac{q+1-i}{M-i} = \frac{q+1}{kq+1-m} T_q. \tag{14}$$

When $q < m$, $T_q = 0$, and with $T_m = 1$. The value of T_q $(q > m)$ can be obtained through Eq. (14). Then, if the value is not an integer, $T'_q = 1$ is defined, and

$$T'_{q+1} = T'_q + \lceil T_{q+1} - T_q \rceil. \tag{15}$$

Thence, the growth function is defined as

$$g(k) = \min\{q : T'_q \geq k\}, \tag{16}$$

where k is the k'th sample. The sampling subset can be expanded according to Eq. (16).

When the probability that the correspondences are by chance inliers to an arbitrary incorrect model is small than η, the iteration process will end. Besides, when the number of iterations comes to the maximum, the process will end too. η is typically set to 95% or 99%, and β indicates the inlier ratio. In estimating a homographic matrix, the probability of an all-inliers sample is approximately equal to β^4:

$$(1-\beta^4)^k < 1-\eta, \tag{17}$$

where k is the iteration number. Equation (17) could be further transformed to

$$k > t = \frac{\log(1-\eta)}{\log(1-\beta^4)}. \tag{18}$$

When k satisfies Eq. (18), it is supposed that an all-inliers sample has already occurred. Figure 2 exhibits the curve of t with respect to β when $\eta_1 = 0.95$ and $\eta_2 = 0.99$. When the inlier ratio is lower than 20%, Fig. 2 reveals that the iteration number increases severely as the inlier ratio decreases.

3.3 *Image Matching via the Local Neighborhood*

This section introduces the proposed method, which is an image matching method via the local neighborhood. Due to the motion slow and smoothness constraints, many matches are likely to be inliers in a small neighborhood of a true match. Thence, it becomes more interested in analyzing the local neighborhood distribution of the feature points. Figure 3 shows an example of a graffiti image and ORB feature points distribution after GMS preprocessing. Figure 3(a) shows a graffiti image with 20×20 grids. Figure 3(b) shows ORB feature points distribution of the red box in image Fig. 3(a).

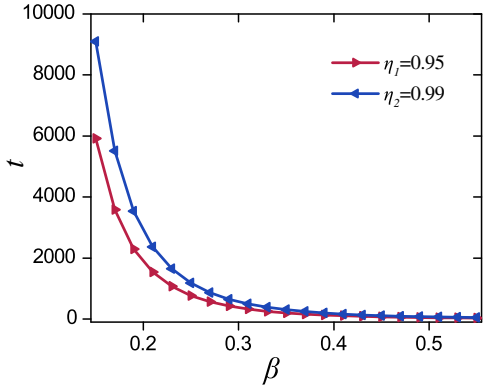

Fig. 2 The curve of t with respect to β, when $\eta_1 = 0.95$ and $\eta_2 = 0.99$. β indicates the inlier ratio. t is the iteration number in which an all-inliers sample is found.

Fig. 3 An example of a graffiti image and ORB feature points distribution after GMS's preprocessing. (a) A graffiti image with 20×20 grids. (b) ORB feature points distribution of the red box in the image (a) after GMS preprocessing.

In Fig. 3(b), there are abundant feature points in the local neighborhood of the feature point named d_i. Conversely, there is just a small amount of feature points in the local neighborhood of the feature point named d_j. Through the motion slow and smoothness constraints, d_i are more likely to be a true match than d_j. Assume that it is subject to normal distribution $N(0, \delta)$, which is the distribution of mutual influence between feature points:

$$G(\Delta x, \Delta y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(\Delta x^2 + \Delta y^2)/2\sigma^2}, \tag{19}$$

where Δx is the difference of the abscissa values of the two points, and Δy is the difference of the ordinate value of the two points. An apposite δ is chosen to count the amount of the feature points in the local neighborhood of a certain feature point, which was denoted by w:

$$w_i = |d_i|, \tag{20}$$

where d_i is the i'th feature point. K-D tree[30] is adopted to quick and efficiently calculate the number of feature points. w is as a weight of one correspondence. Hence, our proposed method adds the weight of one correspondence in image matching, combining the GMS and progressive sample consensus. The proposed method is an image matching method via the local neighborhood, outlined in Algorithm 1.

Algorithm 1 Our image matching method.

Input: an image pair $\{A, B\}$, parameters $\alpha, r, \eta, N, \theta$

Output: inliers

1 Detect ORB feature points and calculate their descriptors;

2 Generate a putative set $C = \{(x_i, y_i)\}_{i=1}^{M}$ using brute-force matching;

3 Compute S_{ab} and τ_a using Eqs. (9) and (10);

4 Obtain a reliable tentative set $C_l = \{(x_i, y_i)\}_{i=1}^{L}$ using Eq. (10);

5 Calculate weight w using Eq. (20);

6 Sort the reliable tentative set $C_l = \{(x_i, y_i)\}_{i=1}^{L}$ in descending order basing on w;

7 Semirandom sample, computer model parameters, and verify model using Eqs. (15) and (18);

8 Identify inliers by the reprojection error.

4 Experimental Results

In this section, we first test the influence of search radius. As the inlier ratio is an important factor in image matching, we then test the robustness of our proposed method on nine typical image pairs with low inlier ratios from the Oxford[31] and Heinly[32] datasets and compare it with the other six state-of-the-art methods, such as RANSAC, GC-RANSAC, PROSAC, GMS, LPM, and RFM-SCAN. Furthermore, we test the image matching capability of our proposed method on the Oxford dataset.

In experiments, the open-source toolbox OpenCV 3.3 is engaged to extract 10,000 ORB feature points. Brute-force matching is employed to generate the putative dataset. In our proposed method, RANSAC, GC-RANSAC, and PROSAC, the termination probability is set to 0.01, and the number of iterations is set to 500 to get a good balance between performance and speed. The scale and rotation variables are both set to true in GMS. The reprojection threshold is set to 3 pixels or 2.5 pixels to identify the true match according to the dataset. The other parameters of each algorithm are set according to the authors' suggestions. The experiments are performed on a laptop PC with an Intel Core i5-7200U, 2.5 GHz CPU, and 16 GB of RAM.

4.1 *Experimental Dataset*

The experimental dataset is the Oxford and Heinly datasets. The Oxford dataset is composed of eight categories of 40 image pairs in total, including bikes, trees, graffiti, wall, bark, boat, Leuven, and UBC. The images involve image blur (bikes, trees), viewpoint change (graffiti, wall), scale and rotation (bark, boat), light change (Leuven), and JPEG compression (UBC). The Heinly dataset includes 40 images, which suffer from dense or sparse viewpoint changes, illumination, pure large-scale zoom, or rotation. Several examples are shown in Fig. 4 from the Oxford and Heinly datasets. Meantime, it provides with the ground truth homographic matrices.

4.2 *Evaluation Metrics*

To quantitatively evaluate the performance of our proposed method, we use three metrics, such as precision, recall, and F-score. Precision is defined as the ratio of the identified correct match number and the whole detected match number:

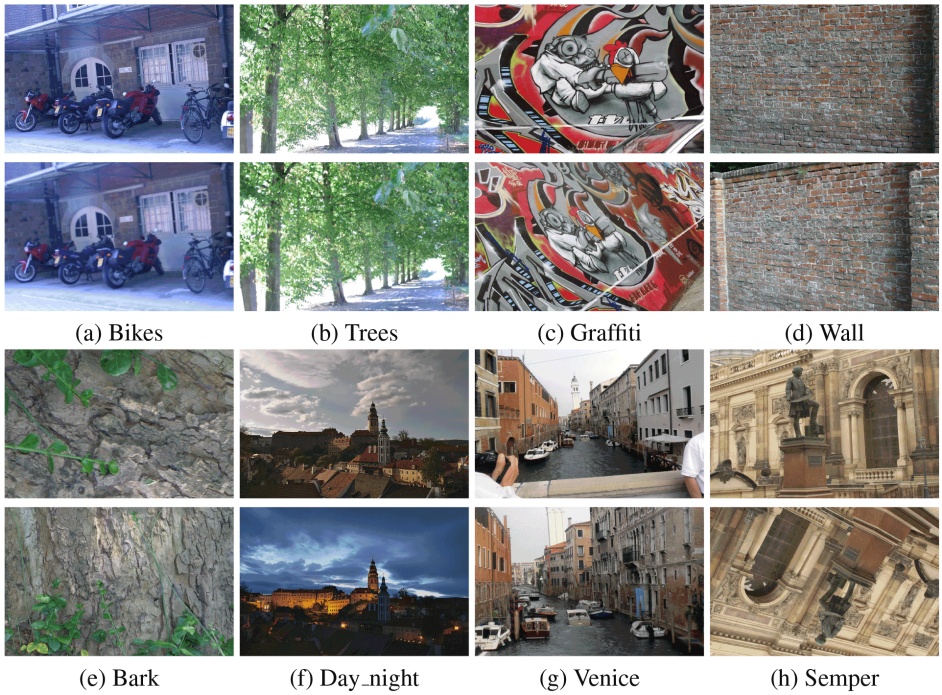

Fig. 4 Examples of image pairs in the Oxford and Heinly datasets.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{21}$$

where TP denotes the identified correct matches, FP denotes the identified incorrect matches. Recall reflects the proportion of the identified correct matches in the ground truth matches:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{22}$$

where FN denotes the correct matches, which are mistaken for the incorrect matches. F-score is a balance that combines the metrics of precision and recall. The F-score is calculated as follows:

$$F - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{23}$$

4.3 *Influence of Search Radius*

In this section, we test the influence of search radius r. Radius r indicates the search range near a certain feature point. If r is small, the search range is limited, even ineffective. If r is large, the search range will increase, which will lead to an increase in running time.

As the images are normalized in our proposed method, we set r with values 0.01, 0.02, and 0.025. Figure 5 reports the inlier ratio, precision, recall, and f-score of our proposed method with these three cases. The average inlier ratio of the Oxford dataset is 40.92%. When r is 0.02 and 0.025, the average precision is 92.39%. When r is 0.01, the average precision is 92.11%. When r is 0.02 and 0.025, the precision is higher than when r is 0.01. When r is 0.025, it has the best average recall (95.01%), followed by when r is 0.02 (93.88%) and when r is 0.01 (93.54%). In

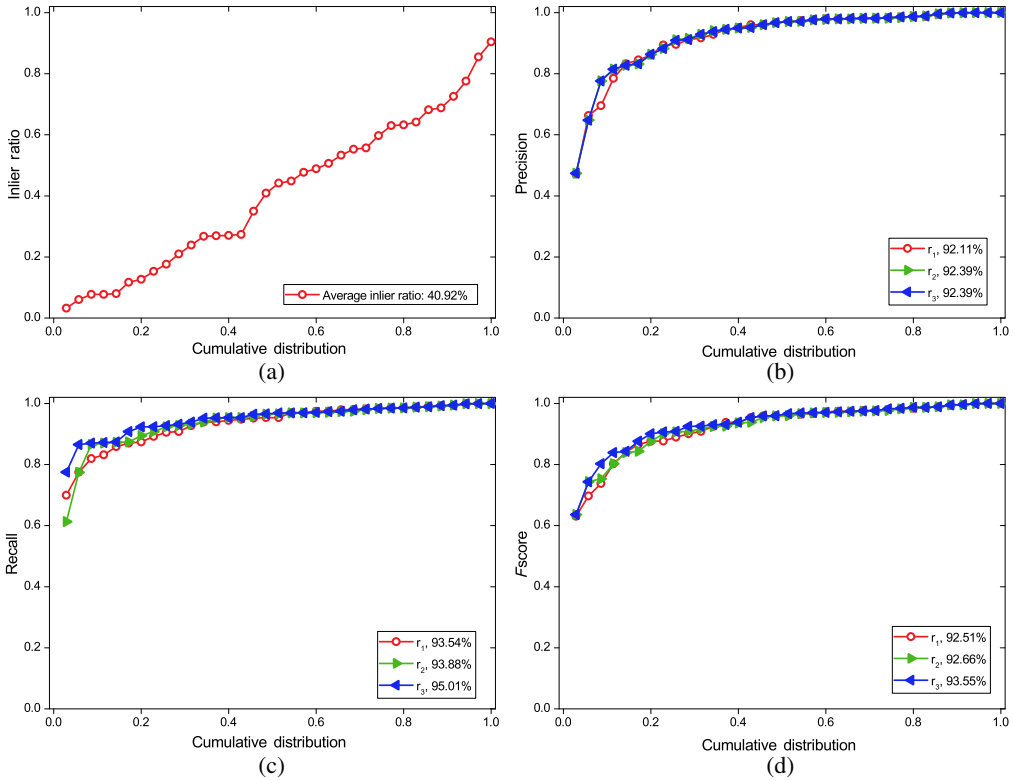

Fig. 5 (a) Inlier ratio, (b) precision, (c) recall, and (d) f-score of our proposed method with respect to the cumulative distribution on the Oxford dataset, which r is set to 0.01, 0.02, and 0.025. The numbers in the boxes represent the average inlier ratio, precision, recall, and f-score.

Table 1 Average running time of our proposed method with three cases (r is set to 0.01, 0.02, and 0.025) on the Oxford dataset. Bold indicates the best result.

	$r = 0.01$	$r = 0.02$	$r = 0.025$
Time (ms)	4.35	4.75	4.86

the performance of f-score, our proposed method with these three cases obtains the same results as the recall.

Table 1 gives the average running time (excluding the cost of ORB feature extraction and GMS) results. When r is 0.01, it has the smallest average running time (4.35 ms), followed by when r is 0.02 (4.75 ms) and when r is 0.025 (4.86 ms). Regarding the running time and a trade-off of the performance of precision, recall, and f-score, r is set to 0.02 in our experiments throughout this study.

4.4 *Robustness Test*

In this section, we test the image matching capability of our proposed method on nine typical image pairs with low inlier ratios from the Oxford and Heinly datasets. We compare our proposed method with six other state-of-the-art methods, such as RANSAC, GC-RANSAC, PROSAC, GMS, LPM, and RFM-SCAN. For fair comparisons, we use the implementations of these algorithms from the open-source websites and set the parameters of each algorithm as reported by the authors. LPM and RFM-SCAN is Matlab code fulfilling in Matlab R2016a. RANSAC, GC-RANSAC, PROSAC, GMS, and our proposed method are c++ code performing in Ubuntu 16.04.

From the Oxford and Heinly datasets, nine typical image pairs are selected, which have low inlier ratios. In the boat category, the inlier ratio of the pair of image1 and image6 is 3.24%. Their inlier ratios are 3.45% (day_night), 6.05% (graffiti), 7.72% (bark), 7.72% (trees), 7.97% (wall), 12.38% (Venice), 22.04% (Semper), and 27.37% (bikes), respectively. We first provide intuitive results of six image matching methods on the bark and wall image pair presented in Figs. 6 and 7, respectively. Figure 6 shows the image matching results of LPM, RFM-SCAN, RANSAC, GMS, GC-RANSAC, and our proposed method on the bark image pair. The bark image pair suffers from scale and rotation. Its inlier ratio is 7.72%. The precision, recall, and f-score of our proposed method are 97.73%, 61.27%, and 75.32%, respectively. Furthermore, PROSAC failed in image matching and was not shown in Fig. 6. In RFM-SCAN, there are many matches of false positives. GC-RANSAC attains the best performance, followed by our proposed method and LPM. Figure 7 exhibits the image matching results of the six image matching methods on the wall image pair. The wall image pair involves viewpoint change. Its inlier ratio is 7.97%. The precision, recall, and f-score of our proposed method are 86.36%, 95.36%, and 90.64%, respectively. PROSAC moreover flunked on the wall image pair and was not shown in Fig. 7. Its reason may be a small number of iterations. Compared with the other five methods, RANSAC has more

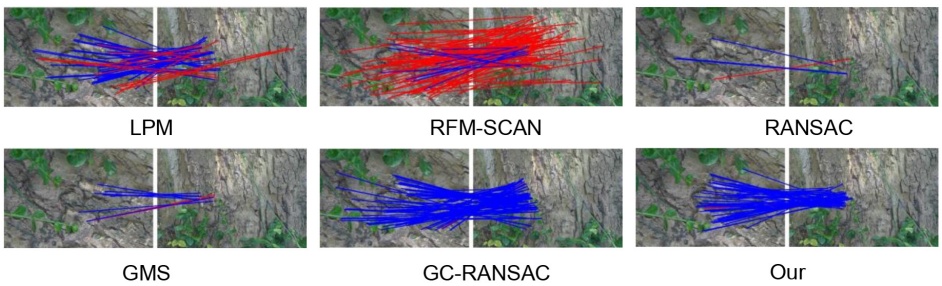

Fig. 6 Image matching results of LPM, RFM-SCAN, RANSAC, GMS, GC-RANSAC, and our proposed method on the bark image pair. Its inlier ratio is 7.72%. The blue line indicates a match of true positive, and the red line indicates a match of false positive. For visibility, at most 100 randomly selected matches are presented in the image pair, and the false negatives are not shown.

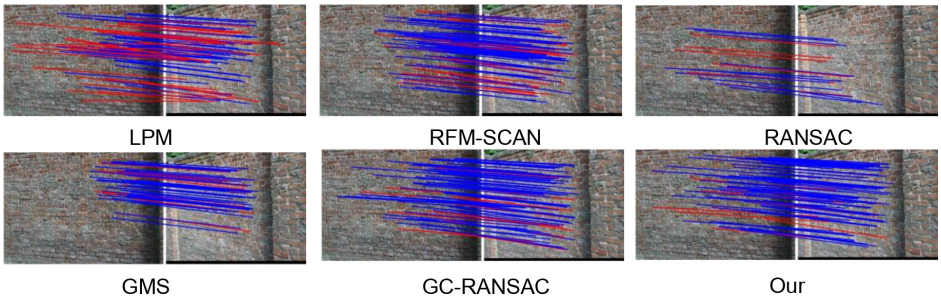

Fig. 7 Image matching results of LPM, RFM-SCAN, RANSAC, GMS, GC-RANSAC, and our proposed method on the wall image pair. Its inlier ratio is 7.97%. The blue line indicates a match of true positive, and the red line indicates a match of false positive. For visibility, at most 100 randomly selected matches are presented in the image pair, and the false negatives are not shown.

matches of false positive on the wall image pair. Figure 7 shows that our proposed method and GC-RANSAC achieve almost the best performance.

Figure 8 reports the curves of inlier ratio, precision, recall, and f-score with respect to the nine selected image pairs, which are the qualitative results of the seven methods. The average inlier ratio of the selected image pairs is 10.88%. PROSAC failed in image matching on seven image pairs. In the putative dataset, the local neighborhood of true match points is contaminated by plenty of false match points. w consequently fails to well reflect the weights of correspondences. Hence, it demonstrates that GMS preprocessing plays a significant role in our proposed method. Our proposed method obtains the best average precision (79.60%), followed by GC-RANSAC (78.05%) and RANSAC (67.21%). Based on the locality preserving constraint,

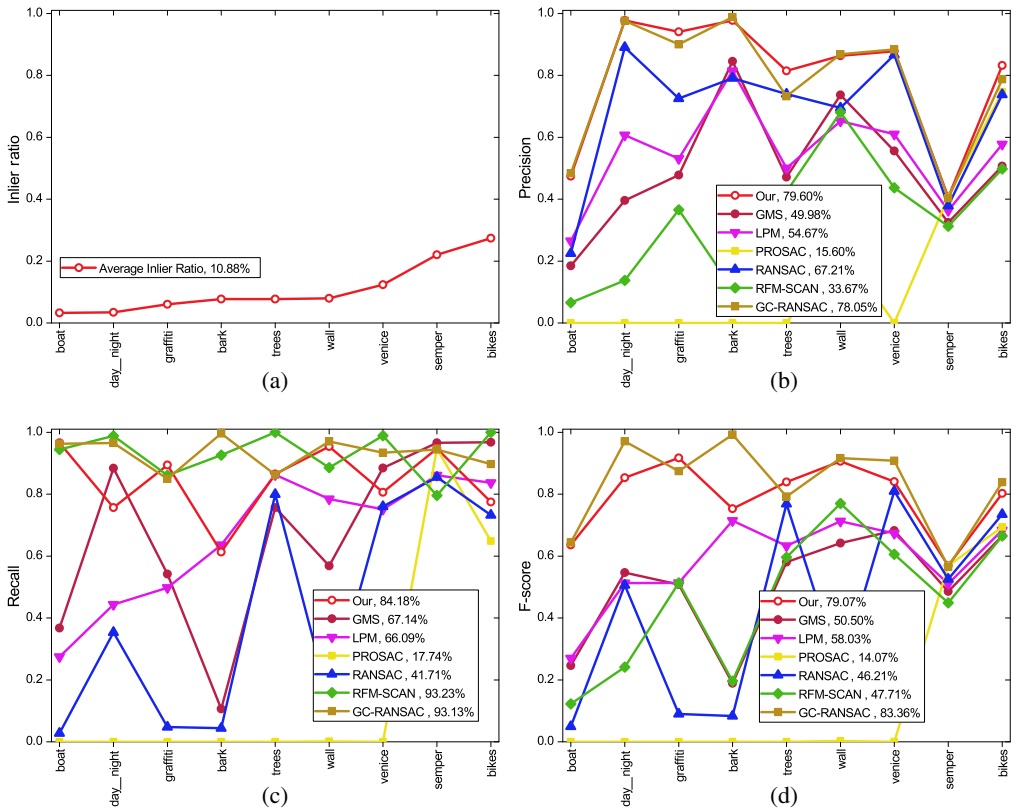

Fig. 8 (a) Inlier ratio, (b) precision, (c) recall, and (d) f-score of RANSAC, GC-RANSAC, PROSAC, GMS, LPM, RFM-SCAN, and our proposed method with respect to the six image pairs selected from the Oxford dataset. The numbers in the boxes represent the average inlier ratio, precision, recall, and f-score.

the average f-score of LPM is 58.03%. GC-RANSAC and our proposed method attain the better average f-score 83.36% and 79.07%, respectively. Figure 8 shows that our proposed method has stronger robustness to establish reliable correspondences than RANSAC, PROSAC, GMS, LPM, and RFM-SCAN.

4.5 *Results on Image Pairs*

In this section, we test the image matching capability of our proposed method on the Oxford dataset and compare it with the other six methods, such as RANSAC, GC-RANSAC, PROSAC, GMS, LPM, and RFM-SCAN. We first provide intuitive results of our proposed method on two typical image pairs as presented in Fig. 9. The interfering type of the graffiti and bikes image pair in Fig. 9 are viewpoint change and blur, respectively. Hence, it is a relative challenge to establish reliable feature matches. Our result is presented on the right side, which illustrates that our proposed method can produce many matches of true positive, few matches of false positive, and false negative. We also present the results of LPM (left) and RANSAC (middle). As shown in Fig. 9, although the inlier ratio of the bikes image pair is higher than the graffiti image pair, the performance of LPM degrades due to various image changes. RANSAC obtains a similar performance to LPM. We can observe that our proposed method achieves the best performance.

Figure 10 shows the curves of inlier ratio, precision, recall, and f-score with respect to the cumulative distribution, which are the image matching results of the seven methods on the Oxford dataset. When the inlier ratio is lower than 5%, RANSAC, GC-RANSAC, GMS, and our proposed method fail. Meanwhile, PROSAC flunked in most cases. In the putative dataset, the local neighborhood of true match points is contaminated by many false match points. w consequently fails to well reflect the weights of correspondences. Furthermore, a small number of iterations is not enough for PROSAC to estimate the deformation. Hence, we do not report these situations. When the inlier ratio is higher than 27.03%, our proposed method, GC-RANSAC, and RANSAC achieve similar precision. When the inlier ratio is lower than 27.03%, our proposed method obtains more precision than the other five methods. GC-RANSAC has the best average precision (92.49%), followed by our proposed method (92.39%) and RANSAC (89.25%). RANSAC gains the fewest average recall (80.10%), due to substandard performance in the putative dataset with an inlier ratio lower than 23.86%. It indicates that RANSAC tends to severely degrade with the decreasing inlier ratio. RFM-SCAN obtains the best average recall (98.02%), followed by GC-RANSAC (97.31%) and our proposed method (93.88%). Similar to the precision performance, when the inlier ratio is higher than 27.03%, our proposed method, GC-RANSAC, and RANSAC achieve a similar f-score. When the inlier ratio is higher than 34.97%, GMS, LPM, and RFM-SCAN obtain a similar f-score. When the inlier ratio is lower than 27.03%, our proposed method and GC-RANSAC obtain better f-score than the other four methods. We observe that our proposed method outperforms the other five state-of-the-art methods. Figure 10 shows that our proposed method

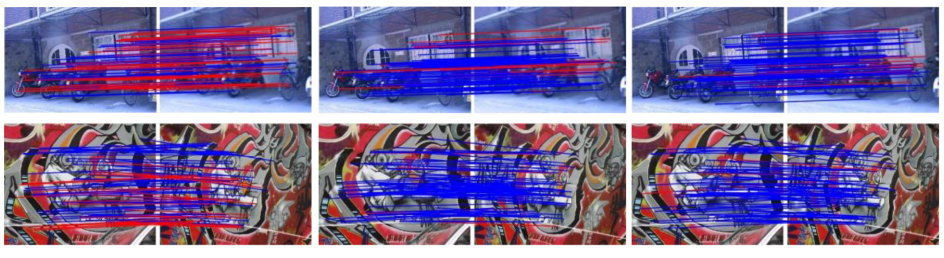

Fig. 9 Image matching results of (a) LPM, (b) RANSAC, and (c) our proposed method on two typical image pairs. The inlier ratio of the graffiti image pair (top row) is 20.94%. The inlier ratio of the bikes image pair (bottom row) is 27.37%. The blue line indicates a match of true positive, the red line indicates a match of false positive. For visibility, at most 100 randomly selected matches are presented in the image pair, and the false negatives are not shown.

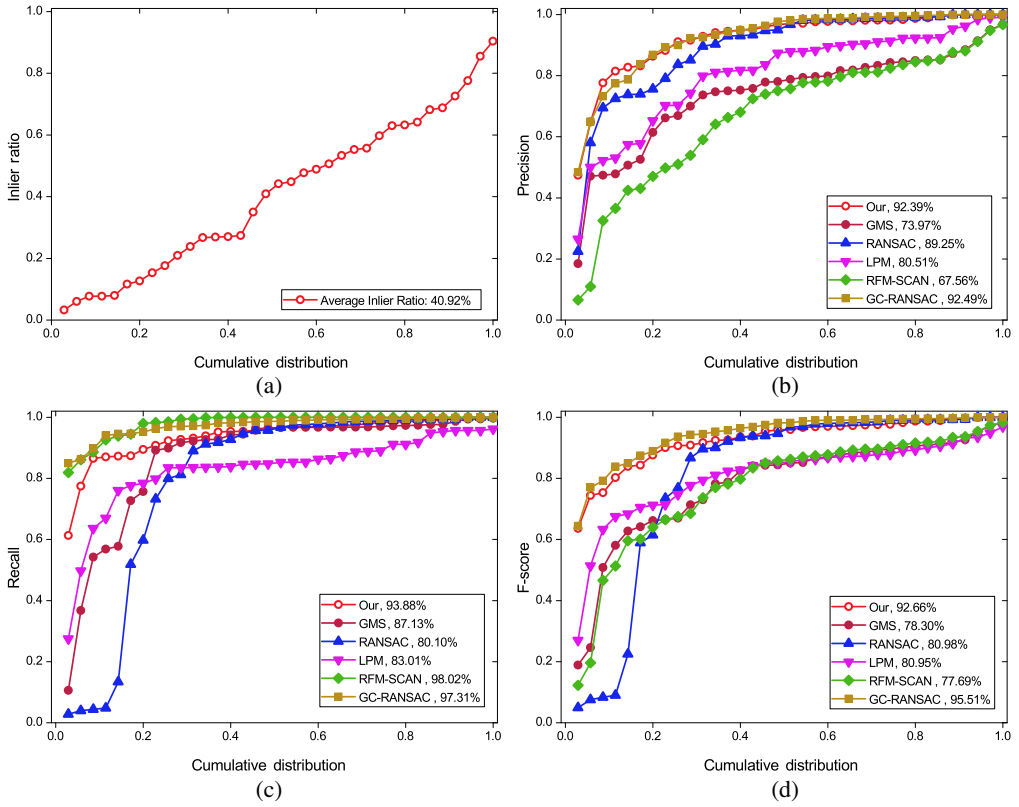

Fig. 10 (a) Inlier ratio, (b) precision, (c) recall, and (d) *f*-score of RANSAC, GC-RANSAC, GMS, LPM, RFM-SCAN, and our proposed method with respect to the cumulative distribution on the Oxford dataset. The numbers in the boxes represent the average inlier ratio, precision, recall, and *f*-score.

Table 2 Average running time of the six methods on the Oxford dataset. Bold indicates the best result.

	RANSAC	GC-RANSAC	PROSAC	GMS	LPM	RFM-SCAN	Our
Time (s)	0.021	0.310	0.013	0.252	0.163	15.752	0.267

has the strongest robustness to establish reliable correspondences than RANSAC, PROSAC, GMS, LPM, and RFM-SCAN, especially in which the inlier ratio is lower than 25%.

Table 2 presents the average running time (excluding the cost of ORB feature extraction) results. RANSAC, GC-RANSAC, PROSAC, GMS, LPM, RFM-SCAN, and our proposed method acquires 0.021, 0.310, 0.013, 0.252, 0.163, 15.752, and 0.267 s, respectively. Table 2 shows that GC-RANSAC takes the most time compared to RANSAC, PROSAC, and our proposed method. Although our proposed method combines the GMS and progressive sample consensus, the average running time of our proposed method is less than the sum of RANSAC and GMS. It indicates that it decreases the running time though engaging the progressive sample consensus.

5 Conclusion

This paper reports an image matching method via the local neighborhood. A key characteristic of our approach is that it is well robust to establish reliable correspondences with a low inlier ratio, especially in which the inlier ratio is lower than 25%. The GMS are engaged to increase the inlier proportion of the putative set. The local neighborhood distribution of feature points is then

gathered as the weights of matches. The progressive sample consensus is employed to estimate a global deformation for removing mismatches. Robust experiments on nine typical image pairs with low inlier ratios illustrate the superiority of our proposed method over RANSAC, PROSAC, GMS, LPM, and RFM-SCAN. The comparison experiments on the Oxford dataset demonstrate that our proposed method outperforms the other five image matching methods.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 51965014).

References

1. X. Li et al., "Underwater image restoration based on modified color-line model," *J. Electron. Imaging* **30**(2), 023010 (2021).
2. X. Wang, Q. Guo, and X. Zhao, "Deep multiscale divergence hashing for image retrieval," *J. Electron. Imaging* **30**(2), 023011 (2021).
3. Y. Yao et al., "Quasi-Newton solver for robust non-rigid registration," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 7600–7609 (2020).
4. J. Li et al., "Robust feature matching via support-line voting and affine-invariant ratios," *ISPRS J. Photogramm. Remote Sens.* **132**, 61–76 (2017).
5. J. Ma et al., "Nonrigid point set registration with robust transformation learning under manifold regularization," *IEEE Trans. Neural Networks Learn. Syst.* **30**(12), 3584–3597 (2019).
6. Y. Lin, Z. Lin, and H. Zha, "The shape interaction matrix-based affine invariant mismatch removal for partial-duplicate image search," *IEEE Trans. Image Process.* **26**(2), 561–573 (2017).
7. M. Awrangjeb and G. Lu, "Techniques for efficient and effective transformed image identification," *J. Vis. Commun. Image Represent.* **20**(8), 511–520 (2009).
8. J. Li, Q. Hu, and M. Ai, "RIFT: multi-modal image matching based on radiation-variation insensitive feature transform," *IEEE Trans. Image Process.* **29**, 3296–3310 (2020).
9. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.* **60**(2), 91–110 (2004).
10. M. Calonder et al., "BRIEF: binary robust independent elementary features," *Lect. Notes Comput. Sci.* **6314**, 778–792 (2010).
11. E. Rublee et al., "ORB: an efficient alternative to SIFT or SURF," in *Int. Conf. Comput. Vision*, IEEE, pp. 2564–2571 (2011).
12. M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM* **24**(6), 381–395 (1981).
13. J. Bian et al., "GMS: grid-based motion statistics for fast, ultra-robust feature correspondence," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 4181–4190 (2017).
14. J. Ma et al., "Locality preserving matching," in *Proc. Twenty-Sixth Int. Joint Conf. Artif. Intell.*, pp. 4492–4498 (2017).
15. J. Li, Q. Hu, and M. Ai, "LAM: locality affine-invariant feature matching," *ISPRS J. Photogramm. Remote Sens.* **154**, 28–40 (2019).
16. X. Jiang et al., "Robust feature matching using spatial clustering with heavy outliers," *IEEE Trans. Image Process.* **29**, 736–746 (2020).
17. M. Fotouhi et al., "SC-RANSAC: spatial consistency on RANSAC," *Multimedia Tools Appl.* **78**(7), 9429–9461 (2019).
18. O. Chum and J. Matas, "Matching with PROSAC-progressive sample consensus," in *IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognit.*, IEEE, Vol. 1, pp. 220–226 (2005).
19. V. Fragoso et al., "EVSAC: accelerating hypotheses generation by modeling matching scores with extreme value theory," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 2472–2479 (2013).
20. O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," *Lect. Notes Comput. Sci.* **2781**, 236–243 (2003).

21. D. Barath and J. Matas, "Graph-cut RANSAC," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 6733–6741 (2018).
22. D. Barath, J. Matas, and J. Noskova, "MAGSAC: marginalizing sample consensus," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 10197–10205 (2019).
23. Q. Tang et al., "A GMS-guided approach for 2D feature correspondence selection," *IEEE Access* **8**, 36919–36929 (2020).
24. G. Wang et al., "Two-view geometry estimation using RANSAC with locality preserving constraint," *IEEE Access* **8**, 7267–7279 (2020).
25. J. Ma et al., "Robust point matching via vector field consensus," *IEEE Trans. Image Process.* **23**(4), 1706–1721 (2014).
26. K. M. Yi et al., "Learning to find good correspondences," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 2666–2674 (2018).
27. C. Zhao et al., "NM-Net: mining reliable neighbors for robust feature correspondences," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 215–224 (2019).
28. J. Ma et al., "LMR: learning a two-class classifier for mismatch removal," *IEEE Trans. Image Process.* **28**(8), 4045–4059 (2019).
29. P.-E. Sarlin et al., "Superglue: learning feature matching with graph neural networks," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit.*, pp. 4938–4947 (2020).
30. M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *Proc. Fourth Int. Conf. Comput. Vision Theory and Appl.*, Vol. 1, pp. 331–340 (2009).
31. K. Mikolajczyk et al., "A comparison of affine region detectors," *Int. J. Comput. Vis.* **65**(1), 43–72 (2005).
32. J. Heinly, E. Dunn, and J.-M. Frahm, "Comparative evaluation of binary features," *Lect. Notes Comput. Sci.* **7573**, 759–773 (2012).

Weiqing Wang is pursuing his PhD in navigation, guidance, and control of Nanjing University of Aeronautics and Astronautics. He received his MSc degree in detection technology and automation from China University of Geosciences, Wuhan, China, in 2009. His research interests include navigation, computer vision, and image processing.

Yongrong Sun is a professor in the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China. He received his PhD in navigation, guidance, and control from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2004. His research interests include inertial navigation, computer vision, image processing, and deep learning.

Zhong Liu is a professor in the School of Mechanical and Electrical Engineering, Guilin University of Aerospace Technology, Guilin, China. He received his PhD in mechanical design and theory from Central South University, Changsha, China, in 2002. His research interests include mechanical design, measuring, image processing, and deep learning.

Zhantian Qin is an associate professor in the School of Mechanical and Electrical Engineering, Guilin University of Aerospace Technology, Guilin, China. He received his MSc degree in mechanical manufacturing and automation from Tianjin University, Tianjin, China, in 2007. His research interests include mechanical design, measuring, and image processing.

Can Wang is an associate professor in the School of Mechanical and Electrical Engineering, Guilin University of Aerospace Technology, Guilin, China. She received her MSc degree in mechanical design and theory from Xiangtan University, Xiangtan, China, in 2006. Her research interests include mechanical design, testing, and image processing.

Jinchang Qin is pursuing his PhD in measuring, testing, and instruments of Nanjing University of Aeronautics and Astronautics. He received his MSc degree in mechanical and electrical engineering from Sichuan University, Chengdu, China, in 2008. His research interests include measuring, image processing, and deep learning.