Pattern Recognition 119 (2021) 108062

Contents lists available at ScienceDirect

Pattern Recognition

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

Mining consistent correspondences using co-occurrence statistics

Guobao Xiao[a], Shiping Wang[b], Han Wang[c], Jiayi Ma[d,*]

[a] Electronic Information and Control Engineering Research Center of Fujian, College of Computer and Control Engineering, Minjiang University, Fuzhou, 350108, China
[b] College of Mathematics and Computer Science, Fuzhou University, Fuzhou, 350108, China
[c] Nanyang Technological University, 639798, Singapore
[d] Electronic Information School, Wuhan University, Wuhan, 430072, China

ARTICLE INFO

Article history:
Received 25 September 2020
Revised 6 March 2021
Accepted 18 May 2021
Available online 27 May 2021

Keywords:
Feature matching
Geometric model fitting
Co-occurrence statistics
Guided sampling

ABSTRACT

In this paper, we propose a mismatch removal method, which mines consistent image feature correspondences using co-occurrence statistics. The proposed method relies on a co-occurrence matrix that counts the number of pixel value pairs co-occurring within the images. Specifically, we propose to integrate the co-occurrence statistics with local spatial information, to preserve the consensus of neighborhood elements. Then, a new measure based on co-occurrence statistics is defined for correspondence similarity, to preserve the consensus of neighborhood topology. After that, with the consensus of neighborhood elements and neighborhood topology, the mismatch removal problem is formulated into a mathematical model, which has a closed-form solution. Extensive experiments show that the proposed method is able to achieve superior or competitive performance on matching accuracy over several state-of-the-art competing methods. In addition, we further exploit the consensus of neighborhood elements and neighborhood topology to propose a novel guided sampling method, which can significantly improve the quality of sampling minimal subsets over state-of-the-arts for two-view geometric model fitting.

1. Introduction

Mining consistent feature correspondences is a fundamental research topic in computer vision tasks, e.g., SLAM, 3D reconstruction, image registration and stereo matching [1–3]. Generally, mining consistent feature correspondences includes two steps, i.e., correspondence generation and correspondence selection. The first step is usually performed by simply picking out local key-point pairs with similar feature descriptors such as SIFT [4]. However, the putative generated correspondences often contain a number of false matches (also called mismatches, i.e. outliers) besides the true matches (i.e. inliers) due to various problems, e.g., local key-point localization errors and ambiguities of the local descriptors. Thus, correspondence selection plays an important role for boosting the geometrical consistency of matches.

In the past several decades, various mismatch removal methods have been proposed, and most of them impose a geometric constraint to formulate the complex problem as an easier geometrical transformation model estimation problem. Nevertheless, it is still a challenging task to mine consistent feature correspondences due to the high computational complexity and the ambiguities of local descriptors. To address the challenges, some mismatch removal methods, e.g., [5–7], are recently proposed by exploiting reliable neighbors of correspondences. These methods can deal with severe deformation by preserving local topological elements. However, the local topological relationship is typically not fully exploited. For example, LPM [5] only adopts the spatial information to count the intersection of neighbors, and ignores the differences among the elements of neighbors, which will lead to the unavailability of local topological structure.

In this paper, we propose an effective non-parametric mismatch removal method. A key new insight of our work is to integrate co-occurrence statistics to construct significant neighborhoods for each correspondence. Co-occurrence information has long been used to measure similarity between textures due to its effectiveness. Particularly, co-occurrences are recently extended to measure similarity between pixel values [8–10], where a co-occurrence matrix counts the number of times that a pair of pixel values co-occurring in an image. Note that, the co-occurrence statistics can capture textures to some degree, since pixel values are probably a part of textured region if they co-occur frequently in the images. Thus, the similarity measured by co-occurrences has nothing to do with the actual pixel values but the co-occurrence statistics. In this paper, we propose to combine co-occurrence statistics (which collect global statistics of texture information, e.g., RGB values and deep features) and local spatial information (which derives from

* Corresponding author.
E-mail addresses: gbx@mju.edu.cn (G. Xiao), shipingwangphd@163.com (S. Wang), hw@ntu.edu.sg (H. Wang), jyma2010@gmail.com (J. Ma).

https://doi.org/10.1016/j.patcog.2021.108062
0031-3203/

the Euclidean distance), to enhance the performance of preserving the consensus of neighborhood elements for matching problems. Moreover, we also introduce the superpixel segmentation technique to the co-occurrence statistics for reducing the computation complexity. After that, we further exploit the co-occurrence statistics between pixel values, and present a new measure for correspondence similarity, which can help preserve the consensus of neighborhood topology to improve the performance of our mismatch removal algorithm. Subsequently, we formulate the mismatch removal into a mathematical model and provide a closed-form solution.

In addition, we extend the proposed mismatch removal method to the application of model fitting. Based on the above analysis in the proposed mismatch removal method, we find that, if two correspondences share high consensus of neighborhood elements and neighborhood topology, then they have a large probability of belonging to the same model instance. Thus, we further exploit the consensus to propose a novel guided sampling method, which can significantly improve the quality of sampled minimal subsets, for two-view geometric model fitting.

More concretely, we summarize the key contributions of this work as follows:

- We propose an effective non-parametric mismatch removal method, which mines consistent image feature correspondences using co-occurrence statistics. The proposed mismatch removal method involves the information from both the global co-occurrence statistics and the local spatial information, to boost the effectiveness of the consensus of neighborhood elements for feature matching. In addition, we define a new measure based on co-occurrence statistics for correspondence similarity, to preserve the consensus of neighborhood topology.
- We propose a novel guided sampling method based on the consensus of neighborhood elements and neighborhood topology to sample high-quality minimal subsets. The proposed sampling method not only covers all model instances in data but also significantly improves the ratio of all-inlier minimal subsets to all sampled minimal subsets (from 52.21% to 84.16% and from 56.95% to 93.39% for homography and fundamental matrix estimation in public available datasets, i.e., AdelaideH [11] and AdelaideF [11], respectively).
- The qualitative and quantitative experiments show that the proposed mismatch removal algorithm and guided sampling method are able to obtain better results over several state-of-the-art competing methods for feature matching tasks and minimum subsect sampling tasks, respectively.

The rest of the paper is organized as follows. We provide an overview of the related work in Section 2. The proposed mismatch removal method is described in details in Section 3, and we extend the proposed method to the model fitting application in Section 4. Experimental results are given in Section 5. Finally, discussion and conclusion are made in Section 6.

2. Related work

In this section, we firstly introduce some mismatch removal methods highly related to this paper, and then we also introduce some guided sampling methods.

According to the principle during the removal process, existing mismatch removal methods can be roughly categorized into three groups, i.e., learning based methods [6,7,12,13], parametric methods [14–17] and non-parametric methods [5,18–24].

The learning based methods [6,7,12,13] introduce the learning strategy to address the matching problems, due to the great success achieved by deep learning in recent years, e.g., [25–28]. For example, learning to find good correspondence (LFGC) [12] is the first one to construct a deep learning framework for correspondence selection, but it cannot address general matching problems, since it requires a specific parametric transformation model. Neighbors mining network (NM-Net) [6] introduces compatibility-specific neighbors to deep network, while learning for mismatch removal (LMR) [7] proposes to learn a general classifier. ACNe [13] presents a learned attentive context normalization to obtain useful context for feature matching. These methods can address the matching problems, however, just as other data-driven methods, they cannot guarantee their performance on the unseen data that are quite different from the training ones, which restricts the application in real world.

The parametric methods [14–17] assume that the inherent transformation of input data conforms to a parametric model, and require a "hypothesis-and-verify" framework (i.e., sampling minimal subsets for hypothesis generation and verifying hypotheses from the putative hypotheses). Thus, they often obtain good performance when the input data are parametric; while they usually cannot obtain reliable performance if the input data are non-parametric.

The non-parametric methods [5,18–24] do not require the assumption and they can deal with more general matching problems. For example, [18,19] analyze the feature direction and correlation computation by singular value decomposition. Graph shift (GS) [20] and identifying correspondence function (ICF) [21] respectively use the graph matching technique and the SVM regression technique to address the matching problems. Coherence based decision boundaries (CODE) [22] and grid-based motion statistics (GMS) [23,24] respectively introduce the coherence based separability constraints and the motion smoothness constraints. Locality preserving matching (LPM) [5] adopts the local neighborhood support to remove mismatches. Thus, these methods usually can handle the data with severe deformations. Among the non-parametric methods, LPM is able to achieve the best performance due to the effectiveness of local topological elements. However, LPM only considers the spatial neighborhood relationship of feature points while ignoring the differences among the elements of neighbors. This will preclude the strict constraint of LPM to preserve local topological structure. We show an example in Fig. 1(a), from which, we can see that, the spatially k-nearest neighbors are not always the true topological structure. In contrast, we combine co-occurrence statistics (that collect global statistics of texture information) and local spatial information to search for k-nearest neighbors, which can avoid the situation, as shown in Fig. 1(b).

3. The proposed mismatch removal method

In this section, we describe the details of the proposed mismatch removal method. Specifically, we first introduce the statement of the mismatch removal problem, and then review the co-occurrence statistics. After that, we propose to preserve the consensus of neighborhood elements and the consensus of neighborhood topology based on the co-occurrence statistics.

3.1. Problem statement

Given two images (I_x, I_y), a set of n putative feature correspondences $S = \{s_i = (x_i, y_i)\}_{i=1}^{n}$ are detected, where x_i and y_i are the spatial positions of two discrete keypoints. The goal of our method is to mine consistent correspondences for identifying each s_i as an inlier or an outlier.

For a simple rigid transformation of the spatial relationship between two images, we can measure the distance between feature correspondences by a certain distance metric, e.g., Euclidean distance. Then we can formulate the mismatch removal problem as a mathematical model. That is, denoting T the unknown inlier set,

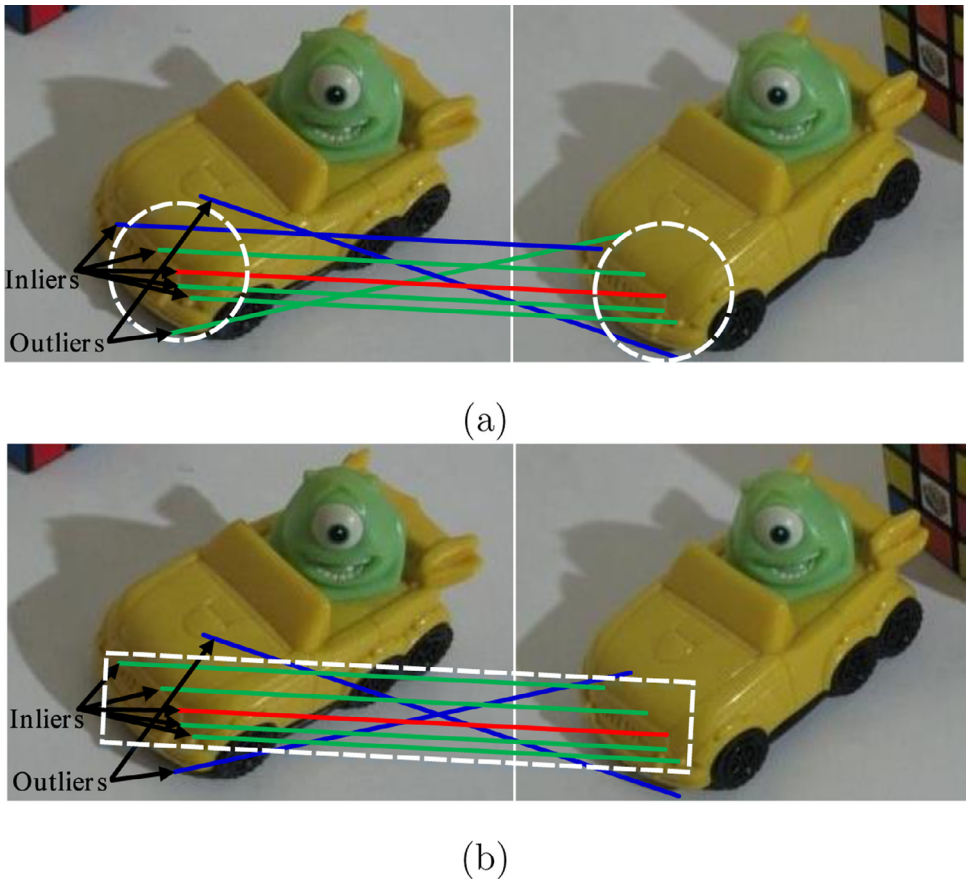

Fig. 1. Visual illustration of (a) spatially k-nearest neighbors without co-occurrence statistics; (b) spatially k-nearest neighbors with co-occurrence statistics. The white box represents the neighbors of a feature point (with red color) in the left image. We also show the inliers and outliers with black arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

its optimal solution is:

$$T^* = \arg\min_T Cf(T; S, \lambda), \tag{1}$$

here, $Cf(T; S, \lambda)$ is the cost function:

$$Cf(T; S, \lambda) = \sum_{i \in T} \sum_{j \in T} \big(d_e(x_i, x_j) - d_e(y_i, y_j)\big)^2 + \lambda(n - |T|), \tag{2}$$

where T^* is the obtained inlier set. d_e is a certain distance metric and λ is a parameter to balance the two terms, i.e., the first term for distance consistency between two feature points, and the second term for outlier penalty. $|\cdot|$ is the cardinality of a set.

The cost function works for a simple rigid transformation, but often fails in a nonrigid transformation while a nonrigid transformation is a frequent phenomenon in real-world tasks. Some methods (e.g., [29]) are proposed to preserve the local neighbor structure of feature points under the nonrigid transformation. Then the cost function in Eq. (2) is redefined as:

$$\begin{aligned} Cf(T; S, \lambda) = \sum_{i \in T} \frac{1}{2M} \Big(& \sum_{j | x_j \in \mathcal{N}_{x_i}} \big(d_e(x_i, x_j) - d_e(y_i, y_j)\big)^2 \\ & + \sum_{j | y_j \in \mathcal{N}_{y_i}} \big(d_e(x_i, x_j) - d_e(y_i, y_j)\big)^2 \Big) \\ & + \lambda(n - |T|), \end{aligned} \tag{3}$$

where $\mathcal{N}_{x_i}$ and $\mathcal{N}_{y_i}$ denote the neighborhood of x_i and y_i, respectively. M is the number of considering neighbors.

To improve the effectiveness for preserving local neighbor structure under severe nonrigid transformation such as scale changes, we replace the absolute distance with a quantized distance. To be specific, denoting $\boldsymbol{t}$ the unknown binary inlier set, where $\boldsymbol{t} = \{t_i\}_{i=1}^n$ and $t_i = 1$ if s_i is an inlier; $t_i = 0$ otherwise, the cost function in Eq. (3) becomes

$$\begin{aligned} Cf(\boldsymbol{t}; S, \lambda) = & \sum_{i=1}^{n} \frac{t_i}{2M} \Big(\sum_{j | y_j \in \mathcal{N}_{y_i}} d_q(x_i, x_j) + \sum_{j | x_j \in \mathcal{N}_{x_i}} d_q(y_i, y_j) \Big) \\ & + \lambda \left(n - \sum_{i=1}^{n} t_i \right), \end{aligned} \tag{4}$$

where $d_q(x_i, x_j)$ is the quantized distance between two feature points with a binary value:

$$d_q(x_i, x_j) = \begin{cases} 0, & x_j \in \mathcal{N}_{x_i}, \\ 1, & \text{otherwise}, \end{cases} \tag{5}$$

and $d_q(y_i, y_j)$ has the same definition as $d_q(x_i, x_j)$.

To improve the generalization of neighborhood representation, we introduce a multi-scale neighborhood representation [5]. That is, we use $\tilde{M} = \{\mathcal{N}_{x_i}^{k_l}\}_{l=1}^m$ to represent the neighborhood of x_i,where m is the scale of neighborhoods and k_l is the size of neighborhoods in each scale. Then the cost function in Eq. (4) becomes

$$\begin{aligned} Cf(\boldsymbol{t}; S, \lambda) = & \sum_{i=1}^{n} \frac{t_i}{2m} \sum_{l=1}^{m} \frac{1}{k_l} \Big(\sum_{j | y_j \in \mathcal{N}_{y_i}^{k_l}} d_q(x_i, x_j) + \sum_{j | x_j \in \mathcal{N}_{x_i}^{k_l}} d_q(y_i, y_j) \Big) \\ & + \lambda \left(n - \sum_{i=1}^{n} t_i \right), \end{aligned} \tag{6}$$

and the quantized distance when $x_j \in \mathcal{N}_{x_i}^{k_l}$ in Eq. (5) is rewritten as

$$d_q(y_i, y_j) = \begin{cases} r@ \quad l0, & y_j \in \mathcal{N}_{y_i}^{k_l}, \\ 1, & \text{otherwise}. \end{cases} \tag{7}$$

Then, the outlier removal problem is formulated as the optimal solution $\boldsymbol{t}$ by minimizing the cost function in Eq. (6).

3.2. Co-occurrence statistics

Given two pixel values (a, b) in an image I, a co-occurrence matrix $Co(a, b)$ is defined as:

$$Co(a, b) = \sum_{p,q} \exp\left(\frac{-d_e(p, q)^2}{2\sigma^2}\right)[I_p = a][I_q = b], \tag{8}$$

where p and q are the corresponding pixel locations, and I_p and I_q are their pixel values in an image I. σ is a user specified parameter and $[\cdot]$ denotes the indicator function.

Intuitively, by Eq. (8), pixel values that frequently co-occur (i.e., inside textured regions) will have a high value in the matrix; Conversely, rarely co-occurring (i.e. across texture boundaries) will have a low value.

Equation (8) can be directly used to deal with a gray scale image, but the co-occurrence matrix will be too large for a color image. Thus, [8] proposed to quantize the RGB values into a small number of color clusters using k-means. However, in this way, k-means still requires to handle all pixels (which are often of a great quantity) in an input image. We, therefore, propose to initially quantize the pixels into a small number (n_s) of superpixels before the clustering process. It is worth pointing out that, for an input image with length n_l and width n_w, the computation complexity of k-means is $O(n_l * n_w * \log(n_l * n_w))$ if we do not use superpixels; while the value is reduced to $O(n_s \log(n_s)) + O(n_l * n_w))$, where $O(n_l * n_w)$ is the computation complexity of a superpixel algorithm, after the superpixel segmentation.

After the two steps for quantizing, a co-occurrence matrix $Co(a, b)$ can then be written as:

$$Co(a, b) = \sum_{p,q} \exp\left(\frac{-d_e(p, q)^2}{2\sigma^2}\right)[T_p = a][T_q = b], \tag{9}$$

where T_p and T_q represent two color clusters.

3.3. Consensus of neighborhood elements

The consensus of neighborhood elements denotes that, if a correspondence s_i is an inlier, then the distributions of neighborhood elements of the two feature points x_i and y_i are similar; Otherwise, the distributions are significantly different. Here, the distribution of neighborhood elements of a feature point is the rank list of neighborhood elements under Euclidean distance.

Note that, we use Eq. (6) to describe the consensus of neighborhood elements in a mathematical model. Specifically, from Eqs. (6) and (7), denoting n_i the number of common neighbors of x_i and y_i, the item $\sum_{j|x_j \in \mathcal{N}_{x_i}^{k_l}} d(y_i, y_j)$ is considered as:

$$\begin{aligned} \sum_{j|x_j \in \mathcal{N}_{x_i}^{k_l}} d_q(y_i, y_j) &= \sum_{j|x_j \in \mathcal{N}_{x_i}^{k_l}, y_j \in \mathcal{N}_{y_i}^{k_l}} d_q(y_i, y_j) \\ &\quad + \sum_{j|x_j \in \mathcal{N}_{x_i}^{k_l}, y_j \notin \mathcal{N}_{y_i}^{k_l}} d_q(y_i, y_j) \\ &= 0 + \sum_{j|x_j \in \mathcal{N}_{x_i}^{k_l}, y_j \notin \mathcal{N}_{y_i}^{k_l}} d_q(y_i, y_j) \\ &= k_l - n_i. \end{aligned} \tag{10}$$

Similarly, $\sum_{j|y_j \in \mathcal{N}_{y_i}^{k_l}} d_q(x_i, x_j) = k_l - n_i$. Then, we can substitute $\sum_{j|y_j \in \mathcal{N}_{y_i}^{k_l}} d_q(x_i, x_j) = \sum_{j|x_j \in \mathcal{N}_{x_i}^{k_l}} d(y_i, y_j)$ into Eq. (6), and obtain the following cost function:

$$\begin{aligned} Cf(\boldsymbol{t}; S, \lambda) = &\sum_{i=1}^{n} \frac{t_i}{m} \sum_{l=1}^{m} \frac{1}{k_l} \sum_{j|x_j \in \mathcal{N}_{x_i}^{k_l}} d_q(y_i, y_j) \\ &+ \lambda(n - \sum_{i=1}^{n} t_i), \end{aligned} \tag{11}$$

Thus, by Eq. (11), if two feature points of a correspondence s_i share more common neighbors, and then s_i has a higher probability of being an inlier; Otherwise, s_i has a higher probability of being an outlier.

From the above analysis, we can see that, the neighborhood construction plays an important role in our method. In this paper, we propose to integrate co-occurrence statistics and local spatial information, instead of only considering the latter, to search for neighborhood elements of each feature point.

Firstly, we adopt co-occurrence statistics (computed by Eq. (9)) and local spatial information (computed by Euclidean distance in a Gaussian model) to measure the correlation between two feature points x_i and x_j. Formally:

$$w(x_i, x_j) = \left(\exp \frac{-d_e(x_i, x_j)^2}{2\sigma_w^2}\right) * Co(a_{x_i}, a_{x_j}), \tag{12}$$

where $d_e(x_i, x_j)$ is the Euclidean distance between x_i and x_j; σ_w is a user specified parameter; a_{x_i} and a_{x_j} are the corresponding pixel values of x_i and x_j, respectively.

From Eq. (12), we can see that, if two feature points x_i and x_j are close in spatial position and their pixel values co-occur frequently, the corresponding correlation value $w(x_i, x_j)$ will be assigned to a large value; Otherwise, it will be assigned a small value. Therefore, we select the feature points with top-k $\{w(x_i, x_j)\}_{j=1}^{k}$ as the neighbors of x_i.

From Fig. 1, which shows a visual illustration of spatially k-nearest neighbors with and without co-occurrence statistics, we can see that, only using spatially local information may lead to selecting the feature points from backgrounds as a neighbor. In contrast, co-occurrence statistics can help avoid this situation. Thus, integrating co-occurrence statistics and local spatial information can effectively preserve the consensus of neighborhood elements.

3.4. Consensus of neighborhood topology

In this subsection, we further design a cost to exploit the consensus of neighborhood topology based on co-occurrence statistics. In the consensus of neighborhood elements, we respectively obtain neighbors of feature points in two images. However, the consensus of neighborhood topology between two images also plays an important role in mining consistent correspondences.

As shown in Fig. 2, although two feature points of a correspondence share common neighbors, their feature values may be significantly different. Thus, we exploit the neighborhood topology by computing the co-occurrence matrix between two correspondences associated with the common neighbors in two images. Specifically, for two correspondences $s_i = \{x_i, y_i\}$ and $s_j = \{x_j, y_j\}$, we first compute the co-occurrence value $\widehat{Co}(s_i, s_j)$ between s_i and s_j:

$$\widehat{Co}(s_i, s_j) = Co(a_{x_i}, a_{x_j}) + Co(a_{y_i}, a_{y_j}). \tag{13}$$

(a) (b)

Fig. 2. Visual illustration of (a) an inlier with consensus of neighborhood topology and (b) outlier without consensus of neighborhood topology based on co-occurrence statistics. The green lines represent the common neighbors of a correspondence (an inlier with red color and an outlier with black color) in two images. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The larger co-occurrence value denotes higher consensus of neighborhood topology. Then, to fuse this into the cost function, i.e. Eq. (11), we define a quantized distance between s_i and s_j:

$$d_q(s_i, s_j) = \begin{cases} 0, & \widehat{Co}(s_i, s_j) > \psi, \\ 1, & \text{otherwise}, \end{cases} \tag{14}$$

where ψ is a predefined threshold. After that, Eq. (11) is rewritten as:

$$Cf(\boldsymbol{t}; S, \lambda, \psi) = \sum_{i=1}^{n} \frac{t_i}{m} \sum_{l=1}^{m} \frac{1}{k_l} \Big(\sum_{j|x_j \in \mathcal{N}_{x_i}^{k_l}} d_q(y_i, y_j) + \sum_{j|x_j \in \mathcal{N}_{x_i}^{k_l}, y_j \in \mathcal{N}_{y_i}^{k_l}} d_q(s_i, s_j) \Big) + \lambda (n - \sum_{i=1}^{n} t_i). \tag{15}$$

3.5. A closed-form solution

In this subsection, we minimize Eq. (15) to obtain the optimal solution. Specifically, we first rewrite the Eq. (15) by merging the terms related to t_i:

$$Cf(\boldsymbol{t}; S, \lambda, \psi) = \sum_{i=1}^{n} t_i (c_i - \lambda) + \lambda n, \tag{16}$$

where

$$c_i = \sum_{l=1}^{m} \frac{1}{m k_l} \Big(\sum_{j|x_j \in \mathcal{N}_{x_i}^{k_l}} d_q(y_i, y_j) + \sum_{j|x_j \in \mathcal{N}_{x_i}^{k_l}, y_j \in \mathcal{N}_{y_i}^{k_l}} d_q(s_i, s_j) \Big). \tag{17}$$

Here, c_i is used to measure if the correspondence s_i satisfies the geometric constraint of preserving local neighborhood structure. Note that, all variables in Eq. (16) are known since we can directly compute the value of c_i for given input data. Intuitively, a correspondence with a smaller value of c_i than λ will correspond to a correct match.

Thus, the optimal solution of $\boldsymbol{t}$ can be obtained by the following criterion:

$$t_i = \begin{cases} 1, & c_i \le \lambda, \\ 0, & \text{otherwise}. \end{cases} \tag{18}$$

Accordingly, the optimal inlier set T^* is computed by:

$$T^* = \{i | t_i = 1, i = 1, \cdots, n\}. \tag{19}$$

Algorithm 1 The proposed mismatch removal algorithm.

Input: Putative feature correspondences $\{s_i\}_{i=1}^n$, parameters m, λ, and ψ.

1: Compute a co-occurrence matrix by Eq. (9).
2: Construct neighborhood of each feature point of every s_i (described in Section 3.3).
3: Compute the co-occurrence value between every two correspondences by Eq. (13).
4: Compute the cost $\{c_i\}_{i=1}^n$ by Eq. (17).
5: Obtain inlier set T^* by Eqs. (18) and (19).

Output: Inlier set T^*.

We summarize the whole procedure of our mismatch removal algorithm in Algorithm 1. The proposed method not only considers the global co-occurrence statistics, but also exploits the local spatial information, for the mismatch removal task. It is worth pointing out that, both our proposed method and LPM [5] adopt the spatial information from the local neighborhood support for feature matching. However, our proposed method introduces co-occurrence statistics from texture information to the local spatial information, which will improve the availability of local topological structure for feature matching.

4. Application to geometric model fitting

In this section, we extend the proposed mismatch removal method to the application of geometric model fitting. We first introduce the statement of the geometric model fitting problem, and then propose an effective subset sampling method, which consists of two steps, i.e., minimum subset initialization and refinement.

4.1. Problem statement

Geometric model fitting is one of the most challenging tasks in computer vision. The aim of geometric model fitting is to estimate the parameters of model instances that best explain input data [16,30]. Generally, there are two steps included in a model fitting framework: 1) sample minimal subsets to generate model hypotheses; 2) select model hypotheses to estimate the parameters of model instances.

Thus, sampling minimal subsets is a key step of model fitting. Here, a minimal subset contains the minimum number ($\hat{p}$) of data points that are required to estimate a model hypothesis (e.g., 4 for homography model), and it can be classified into two types: an all-inlier minimal subset which consists of all inliers belonging to

a same model instance, and an impure minimal subset which contains at least one outlier (gross outlier or pseudo outlier[1]). To "hit" all model instances in data[2], the required number of sampled minimal subsets increases exponentially with the outlier ratio and the minimal subset size ($\hat{p}$).

For example, to hit a model instance with the probability β, RANSAC [14] requires to sample $\frac{log(1-\beta)}{log(1-(1-\alpha)^{\hat{p}})}$ minimal subsets at least, where α denotes the outlier ratio. That is, RANSAC requires to sample a large number of minimal subsets when the data include a large proportion of outliers. Accordingly, many guided sampling algorithms [16,31–34] are proposed to improve the quality of minimal subsets.

Existing guided sampling methods can be roughly classified into two categories, i.e., deterministic sampling methods [16,35] and random sampling methods [14,31–34,36,37]. AM-RS [35] formulates the model fitting problem as a global optimization problem. SDF [16] introduces the superpixel segmentation to boost all-inlier minimal subsets. These guided sampling methods can sample many promising minimal subsets but they are often computationally expensive.

For the random sampling methods, e.g., NAPSAC [34], LO-RANSAC [33], G-MLESAC [36], PROSAC [32], Multi-GS [31] and RCMSA [37], they are proposed to reduce the number of sampled minimal subsets while hitting all structures with reasonable time. NAPSAC and LO-RANSAC adopt local information to refine the sampled minimal subsets. G-MLESAC and PROSAC use some information such as matching scores to sample promising minimal subsets. Multi-GS and RCMSA adopt the statistical information from the current sampled minimal subsets to boost all-inlier ones. These sampling methods are able to improve the quality of minimal subsets, but they are still far from being practical to deal with real-world problems due to the low value of ϱ, where ϱ is the ratio of all-inlier minimal subsets to all sampled minimal subsets.

Thus, sampling a pure minimal subset is not an easy task due to the influence of outliers and multiple structures. Recall that, the aim of feature matching is to find consistent feature correspondences. Although the correspondences obtained by feature matching methods may also include some outliers, most of outliers are removed, which can reduce the bad influence of outliers to sample a pure minimal subset. In addition, the consistent neighborhood information used by the proposed mismatch removal method can be further exploited to improve the quality of minimal subsets.

4.2. The proposed guided sampling method

To address the above issue for two-view model fitting problem, we propose to mine consistent feature correspondences (i.e. matches) in advance, and then sample minimal subsets from these feature correspondences. Specifically, based on the consistent feature correspondences obtained by the proposed mismatch removal method, we exploit the consensus of neighborhood elements to initialize some effective minimal subsets, and then refine the sampled minimal subsets according to consensus of neighborhood topology.

4.2.1. Minimum subset initialization

After removing outliers estimated by the proposed mismatch removal algorithm, we sample the minimal subsets from the remaining data points.

Specifically, for each correspondence $s_i = \{x_i, y_i\}$, we first obtain its consistent neighbors, i.e. the correspondences whose feature points in two views are both the neighbors of the two feature points x_i and y_i, respectively. Then, we sample the elements of a minimal subset from the consistent neighbors, according to the co-occurrence value (computed by Eq. (13)). That is, being an element of a minimal subset for a correspondence s_j is proportional to the consensus of neighborhood topology between s_i and s_j.

After that, we remove some invalid minimal subsets whose elements are smaller than the minimal subset size ($\hat{p}$), and finish the minimum subset initialization.

4.2.2. Minimum subset refinement

After the minimum subset initialization, we further refine the minimum subsets. Note that, some minimum subsets consist of inliers, but the elements may belong to different model instances since the input data for model fitting often include multiple model instances.

Therefore, we propose to filter out some correspondences, which is dissimilar to other elements of a minimum subset based on the ratio of length and the angle among them. Specifically, we measure the consensus of a correspondence s_i to a minimum subset $\boldsymbol{s}_j = \{s_i\}_{i=1}^{\hat{n}_j}$ as:

$$Cs(s_i, \boldsymbol{s}_j) = \frac{\min\{|s_i|, |med(\boldsymbol{s}_j)|\}}{\max\{|s_i|, |med(\boldsymbol{s}_j)|\}} \cdot \frac{\left(|s_i|, med(\boldsymbol{s}_j)\right)}{|s_i| \cdot |med(\boldsymbol{s}_j)|}, \tag{20}$$

where $med(\boldsymbol{s}_j)$ is the median value of all elements in $\boldsymbol{s}_j$. $(\cdot,\cdot)$ and $|\cdot|$ denote the inner product and the corresponding induced norm, respectively. Then, a larger value of $Cs(s_i, \hat{\boldsymbol{s}}_j) \in [-1, 1]$ denotes higher consensus of the minimal subset. Thus, we update each minimum subset $\boldsymbol{s}_j$ according to the value of $Cs(s_i, \hat{\boldsymbol{s}}_j)$, i.e.,

$$\hat{\boldsymbol{s}}_j = \{s_i | Cs(s_i, \boldsymbol{s}_j) < \mu, i = 1, 2, ..., \hat{n}_j\}, \tag{21}$$

where μ is a predefined threshold.

After that, we also remove some invalid minimal subsets whose elements are smaller than the minimal subset size ($\hat{p}$), and then finish the minimum subset sampling process.

We also summarize the whole procedure of our guided sampling method in Algorithm 2. It is worth pointing out that the superiority of the guided sampling method over existing methods comes from three aspects. Firstly, we remove most of outliers by the proposed mismatch removal method, which can reduce the influence of outliers on the sampling results. Secondly, we exploit the consensus of neighborhood elements of each data point to sample minimal subsets, which can cover all model instances in data. Thirdly, we adopt the consensus of neighborhood topology to refine the sampled minimal subsets, which can effectively improve the number of all-inlier minimal subsets. Thus, the proposed guided sampling method can achieve more promising sampling performance than that achieved by state-of-the-art methods.

Algorithm 2 The proposed guided sampling method.

Input: Putative feature correspondences $\{s_i\}_{i=1}^{n}$.

1: Estimate inlier set $\boldsymbol{t}^*$ by Algorithm 1.
2: Update the neighborhood of each estimated inlier (described in Section 3.3).
3: Compute the co-occurrence value between every two estimated inliers by Eq. (13).
4: Initialize minimum subsets (described in Section 4.2.1).
5: Filter out some elements of each minimal subset (described in Section 4.2.2).

Output: Minimal subsets.

5. Experimental results

In this section, we compare the proposed mismatch removal algorithm (called COMR) and the proposed sampling method (called

[1] A pseudo outlier is an inlier belonging to another model instance.

[2] Here, hitting all model instances means that at least one all-inlier minimal subset corresponding to each model instance in data is sampled.

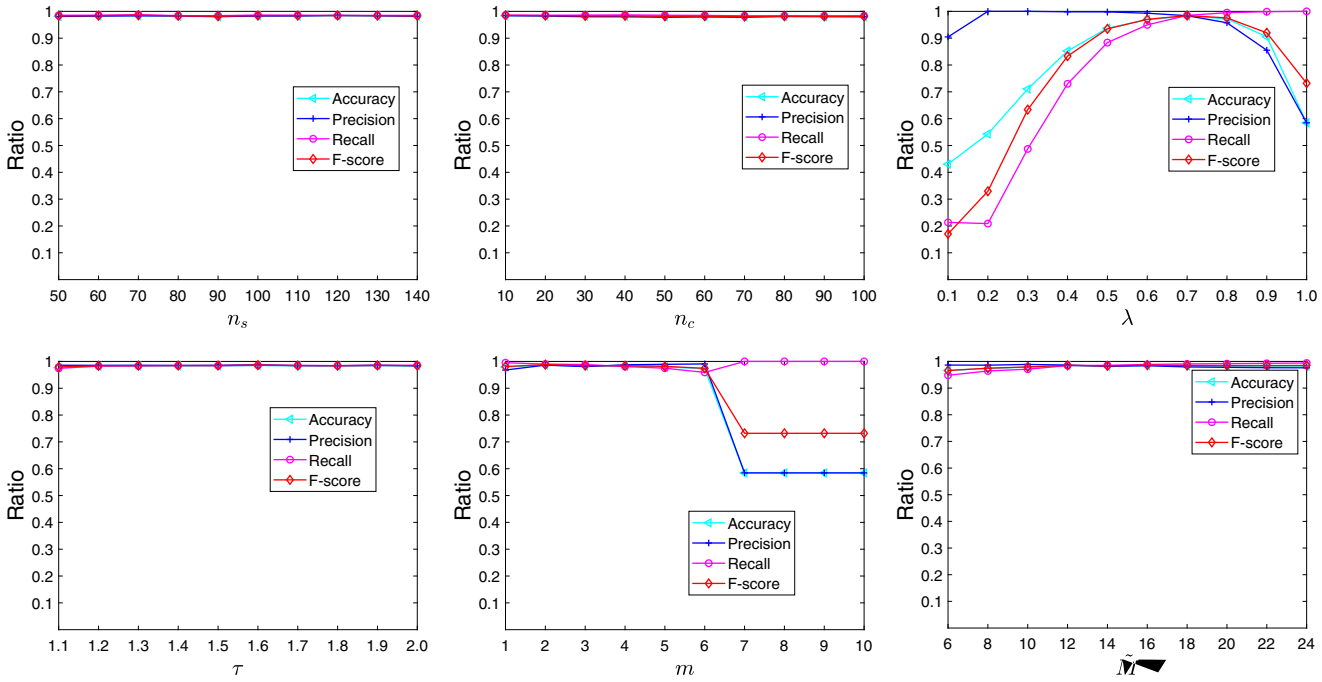

Fig. 3. Quantitative comparison by different settings of parameters for the proposed method.

COS) with several state-of-the-art methods for the feature matching task and minimal subsets sampling task, respectively. All experiments are run on MS Windows 10 with Intel Core i7 − 8565 CPU 1.8*GHz* and 16*GB* RAM.

5.1. Results on feature matching

In this subsection, we compare COMR with several state-of-the-art methods, including GS [20], ICF [21], GMS [23,24] and LPM [5], for feature matching tasks. We choose these representative methods since they are non-parametric methods as COMR. We also run RANSAC [14] as a baseline.

Datasets and evaluation metrics. To provide quantitative comparisons with state-of-the-art competing methods, we conduct experiments on five public available datasets, i.e., VGG [38], DAISY [39], DTU [40], AdelaideH [11] and AdelaideF [11]. VGG contains 40 image pairs obeying homography with single consistency. DAISY contains 52 image pairs with wide baselines. DTU contains 131 image pairs with large viewpoint changes. AdelaideH and AdelaideF respectively contain 19 image pairs obeying homography and fundamental matrix with multiple consistencies.

We use accuracy, precision, recall and F-score to characterize the matching performance, where accuracy is defined as the ratio of the identified match number and the total match number, precision is defined as the ratio of the identified inlier number and the preserved match number, recall is defined as the ratio of the identified inlier number and the total inlier number, and F-score is defined as the ratio of $2 * Precision * Recall$ and $Precision + Recall$.

5.1.1. Parameter analysis

There are six parameters in the proposed COMR: n_s, n_c, λ, ψ, $\tilde{M}$, and m. n_s is the number of superpixels and n_c is the value of k for k-means in Section 3.2. λ is a parameter to balance the two terms in Eq. (15). ψ is a threshold to determine whether two correspondences preserve the consensus of neighborhood topology in Eq. (14). $\tilde{M}$ and m are the number and the scale of the nearest neighbors for multi-scale neighborhood construction in Eq. (15), respectively. We test different settings on AdelaideF, and report accuracy, precision, recall and f-score in Fig. 3.

We can see that, the results are very stable with different values of n_s, n_c, ψ and $\tilde{M}$. Thus, the proposed method is not sensitive to these parameters. For λ, the proposed method is able to obtain the best performance when $\lambda = 0.7$. For m, the proposed method is able to obtain stable results when $2 \leq m \leq 6$. Thus, we set $n_s = 100$, $n_c = 10$, $\psi = 2$, $\tilde{M} = 8$, $\lambda = 0.7$ and $m = 3$.

5.1.2. Algorithm analysis

Note that the proposed COMR consists of two key steps, i.e. preserving the consensus of neighborhood elements and neighborhood topology. For the consensus of neighborhood elements, the main improvement proposed by COMR is to integrate co-occurrence statistics. Thus, to show the importance of co-occurrence statistics, we test two versions, i.e., COMR1 without co-occurrence statistics and COMR2 with co-occurrence statistics. We also test another version, i.e., COMR3 with preserving consensus of neighborhood topology, to show the importance of this part.

Figure 4 shows a quantitative comparison with respect to the cumulative distribution obtained by three versions of COMR. For simplicity, we only use the AdelaideF for evaluation. We can see that, COMR2 can achieve better performance on accuracy, precision and F-score than COMR1, and COMR2 also achieves similar recall values as COMR1. This can show that co-occurrence statistics can help improve the matching performance of the proposed method. Recall that we introduce the consensus of neighborhood topology to the proposed method, and the better performance obtained by COMR3 over COMR1/COMR2 can also validate its importance.

5.1.3. Results on public image datasets

We next provide a quantitative evaluation of our COMR on five public available datasets, and compare to five state-of-the-art methods, i.e., RANSAC, GS, ICF, GMS and LPM. We select eight representative image pairs from the five datasets and present the results in Fig. 5, to provide an intuitive illustration of the feature matching performance of COMR. We also report the initial inlier percentage, accuracy, precision, recall, and F-score statistics of the six methods in Fig. 6.

From the results, we can see that, for VGG, RANSAC is able to achieve good performance on accuracy, precision, recall and F-

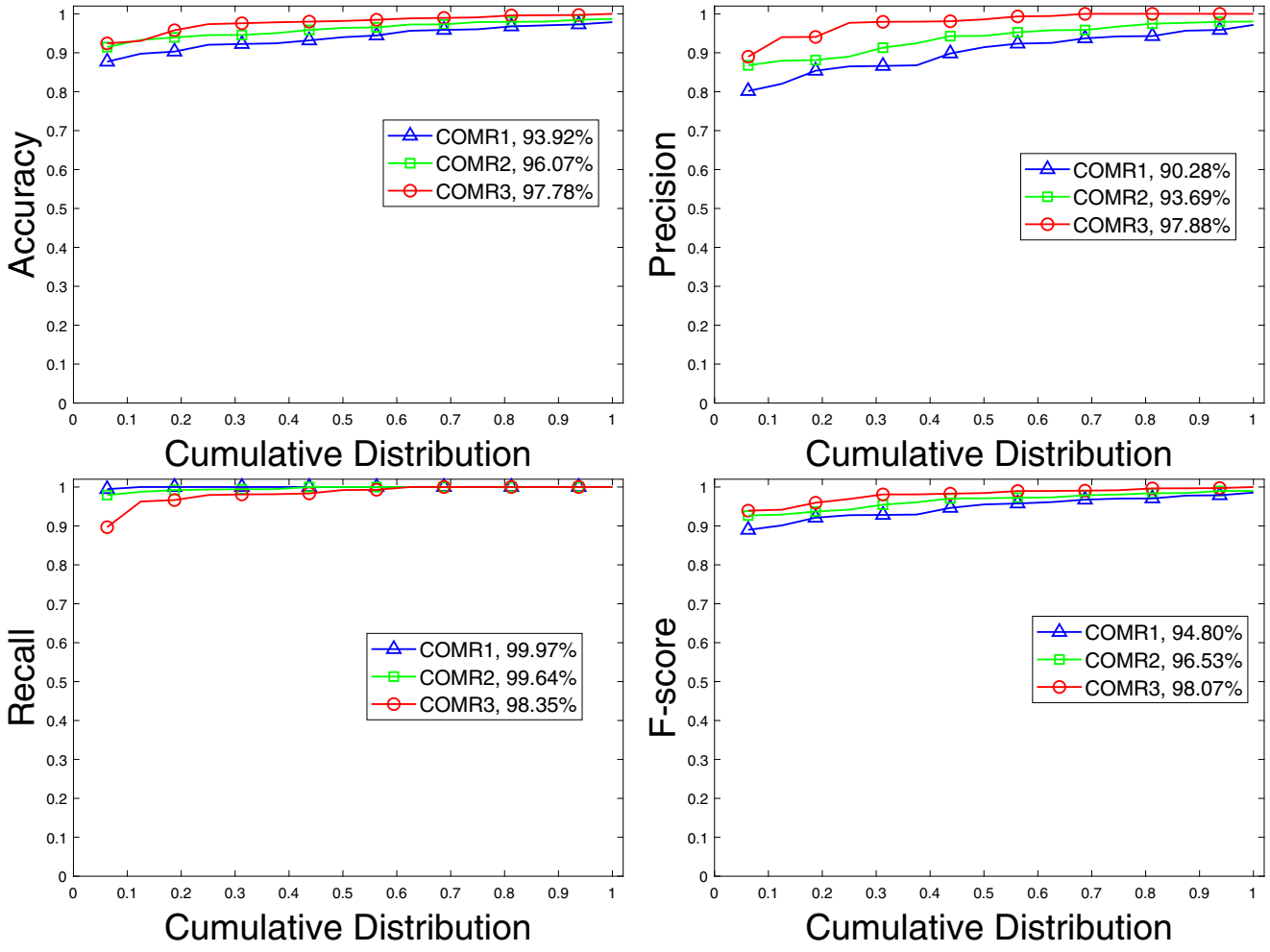

Fig. 4. Quantitative comparison with respect to the cumulative distribution obtained by different versions of COMR. A point on the curve with coordinate (x, y) denotes that there are $100 * x$ percents of image pairs which have values no more than y.

score, especially for F-score (it achieves the largest values among all six competing methods). This is because the image pairs in VGG obey homography with single consistency, and RANSAC requires a geometrical model between image pairs, thus it can work well for this dataset. Note that GS and ICF also require a motion model between image pairs and GMS is sensitive to the scale changes (recall that the image pairs in DAISY and DTU respectively include wide baselines and large viewpoint changes, and the image pairs in AdelaideH and AdelaideF include multiple consistencies). In contrast, LPM and COMR do not have the requirement, and they can achieve good results on different types of datasets.

Compared to LPM, COMR achieves better results on the performance of accuracy, precision and F-score for all five datasets (their recall statistics do not have obvious differences). Although both of them adopts the information of reliable neighbors, COMR integrates co-occurrence statistics to construct neighborhoods for each correspondence, and COMR also introduces co-occurrence statistics to preserve the consensus of neighborhood topology, which can further improve the matching performance.

5.2. Results on minimal subsets sampling

In this subsection, we compare COS with eight state-of-the-art sampling methods, including RANSAC, NAPSAC [34], LO-RANSAC [33], G-MLESAC [36], PROSAC [32], Multi-GS [31], RCMSA [37] and SDF [16] on two popular datasets (i.e. AdelaideH and AdelaideF, for homography and fundamental matrix estimation, respectively). Here, we do not test the sampling methods on the other three datasets (i.e., VGG, DAISY and DTU) for feature matching since VGG only contains single-structure data and its inlier ratio is high and thus all sampling methods can achieve good sampling results. For DAISY and DTU, the datasets do not obey geometric constraints for model fitting. In contrast, AdelaideH and AdelaideF contain multi-structure data and obey different geometric constraints for model fitting. To show the importance of minimum subset filter, we test two versions of COS, i.e. COS1 without minimum subset filter and COS2 with minimum subset filter. Note that, there is a parameter μ in Eq. (21) for COS2. This parameter controls the number of correspondences that are removed by COS2, that is, the larger value of μ the less correspondences are removed. Thus, we empirically set μ as 0.2.

We repeat all competing methods 50 times, and show the mean value statistics of the total number of sampling minimal subsets and the value of ϱ obtained by the ten competing methods within 5 CPU seconds on all image pairs of AdelaideH and AdelaideF in Fig. 7. We also select six image pairs that contain different numbers of model instances for the two model fitting tasks, and report the details, i.e., the total number of sampled minimal subsets, the value of ϱ, the number and the ratio of sampled minimal subsets belonging to different model instances, in Table 1.

From Fig. 7 and Table 1, we can see that COS1/COS2 have significantly improved the value of ϱ over other eight competing meth-

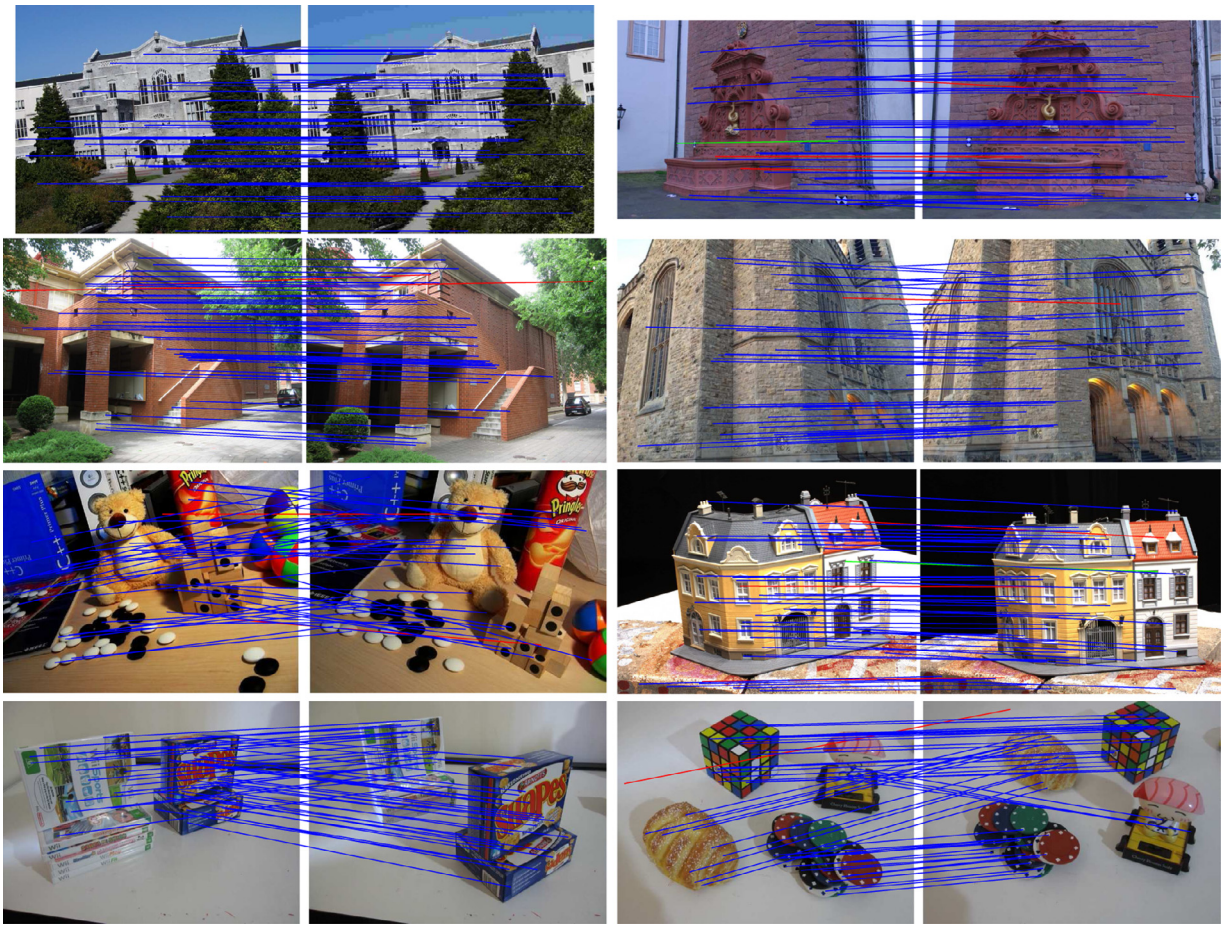

Fig. 5. Qualitative illustration of feature matching performance of our COMR on eight representative image pairs. From left to right and top to bottom: Ubc, Fountain03, Johnsonb, Bonhall, BearB01, House4547, Gamebiscuit and Cubebreadtoychips. For visibility, at most 50 randomly selected matches without true negatives are presented (blue = true positive, green = false negative, red = false positive). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Performance of the ten competing sampling methods for homography (D1-D3) and fundamental matrix estimation (D4-D6). #Total reports the total number of minimal subsets sampled within 5 CPU seconds (and the ratio of all-inlier minimal subset to the total number of subsets); #S-*i* reports the number of all-inlier minimal subset sampled for the *i*th structure (and the corresponding ratios); The best result of each performance measure is boldfaced. D1-Barrsmith; D2-Elderhallb; D3-Unihouse; D4-Breadtoy; D5-toycubecar; D6-Cubebreadtoychips.

Method		RANSAC	NAPSAC	LO-RANSAC	G-MLESAC	PROSAC	Multi-GS	RCMSA	SDF	COS1	COS2
D1	#Total	24452(0.19)	24029(1.61)	**25000**(0.22)	24725(1.97)	**25000**(0.23)	1401(13.35)	1919(12.66)	1264(10.13)	**25000**(67.60)	**25000(86.35)**
	#S1	44(0.18)	293(1.22)	54(0.22)	485(1.96)	53(0.21)	177(12.63)	214(11.15)	104(8.23)	10217(40.87)	**12830(51.32)**
	#S2	3(0.01)	94(0.39)	0(0.00)	3(0.01)	4(0.02)	10(0.71)	29(1.51)	24(1.90)	6683(26.73)	**8757(35.03)**
D2	#Total	24452(0.46)	23902(8.25)	**25000**(0.39)	**25000**(5.77)	**25000**(0.41)	1401(18.63)	1409(8.52)	1638(21.37)	**25000**(60.73)	**25000(67.04)**
	#S1	19(0.08)	626(2.62)	21(0.08)	30(0.12)	10(0.04)	77(5.50)	2(0.14)	133(8.12)	**4420(17.68)**	4418(17.67)
	#S2	5(0.02)	89(0.37)	4(0.02)	29(0.12)	2(0.01)	10(0.71)	0(0.00)	14(0.85)	1913(7.65)	**1978(7.91)**
	#S3	90(0.36)	1256(5.25)	72(0.29)	1383(5.53)	90(0.36)	174(12.42)	118(8.37)	203(12.39)	8850(35.40)	**10363(41.45)**
D3	#Total	1180(1.27)	1029(59.09)	1098(3.37)	1134(4.50)	1134(7.14)	201(27.36)	99(12.12)	437(65.90)	**1759**(88.46)	1756(**88.67**)
	#S1	4(0.34)	171(16.62)	3(0.27)	0(0.00)	0(0.00)	17(8.46)	1(1.01)	114(26.09)	463(26.32)	**464(26.42)**
	#S2	0(0.00)	2(0.19)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	4(**4.04**)	0(0.00)	**26**(1.48)	18(1.03)
	#S3	4(0.34)	162(15.74)	1(0.09)	12(1.06)	19(1.68)	12(5.97)	2(2.02)	89(20.37)	447(25.41)	**452(25.74)**
	#S4	7(0.59)	207(20.12)	33(3.01)	39(3.44)	62(5.47)	26(12.94)	5(5.05)	85(19.45)	470(26.72)	**479(27.28)**
	#S5	0(0.00)	66(6.41)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	**150(8.53)**	144(8.20)
D4	#Total	7895(0.08)	7621(34.86)	7745(0.32)	7726(13.00)	7504(0.84)	831(33.94)	909(46.75)	846(65.25)	**25000(100.00)**	**25000(100.00**)
	#S1	6(0.08)	2159(28.33)	25(0.32)	1004(13.00)	63(0.84)	220(26.47)	227(24.97)	528(62.41)	**17598(70.39)**	17551(70.20)
	#S2	0(0.00)	498(6.53)	0(0.00)	0(0.00)	0(0.00)	62(7.46)	198(21.78)	24(2.84)	7402(29.61)	**7449(29.80)**
D5	#Total	11695(0.01)	11423(17.31)	11788(0.16)	11526(1.36)	11493(0.08)	1152(28.04)	1370(65.04)	1185(79.75)	**25000**(91.43)	**25000(93.87)**
	#S1	1(0.01)	629(5.51)	0(0.00)	0(0.00)	0(0.00)	112(9.72)	423(30.88)	255(21.52)	7578(30.31)	**8270(33.08)**
	#S2	0(0.00)	1348(11.80)	19(0.16)	157(1.36)	9(0.08)	211(18.32)	468(34.16)	690(58.23)	**14998(59.99)**	14752(59.01)
	#S3	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	281(1.12)	**445(1.78)**
D6	#Total	7519(0.01)	7168(14.36)	7220(0.08)	7389(0.11)	7217(0.01)	781(39.82)	789(41.44)	840(43.33)	**25000**(96.27)	**25000(96.36**)
	#S1	0(0.00)	507(7.07)	1(0.01)	1(0.01)	0(0.00)	136(17.41)	6(0.76)	148(17.62)	**7818(31.27)**	7794(31.18)
	#S2	0(0.00)	190(2.65)	0(0.00)	0(0.00)	0(0.00)	45(5.76)	156(19.77)	52(6.19)	4824(19.30)	**4918(19.67)**
	#S3	0(0.00)	58(0.81)	0(0.00)	0(0.00)	0(0.00)	19(2.43)	110(13.94)	28(3.33)	**4058(16.23)**	3940(15.76)
	#S4	1(0.01)	274(3.82)	5(0.07)	7(0.09)	1(0.01)	111(14.21)	55(6.97)	136(16.19)	7368(29.47)	**7439(29.76)**

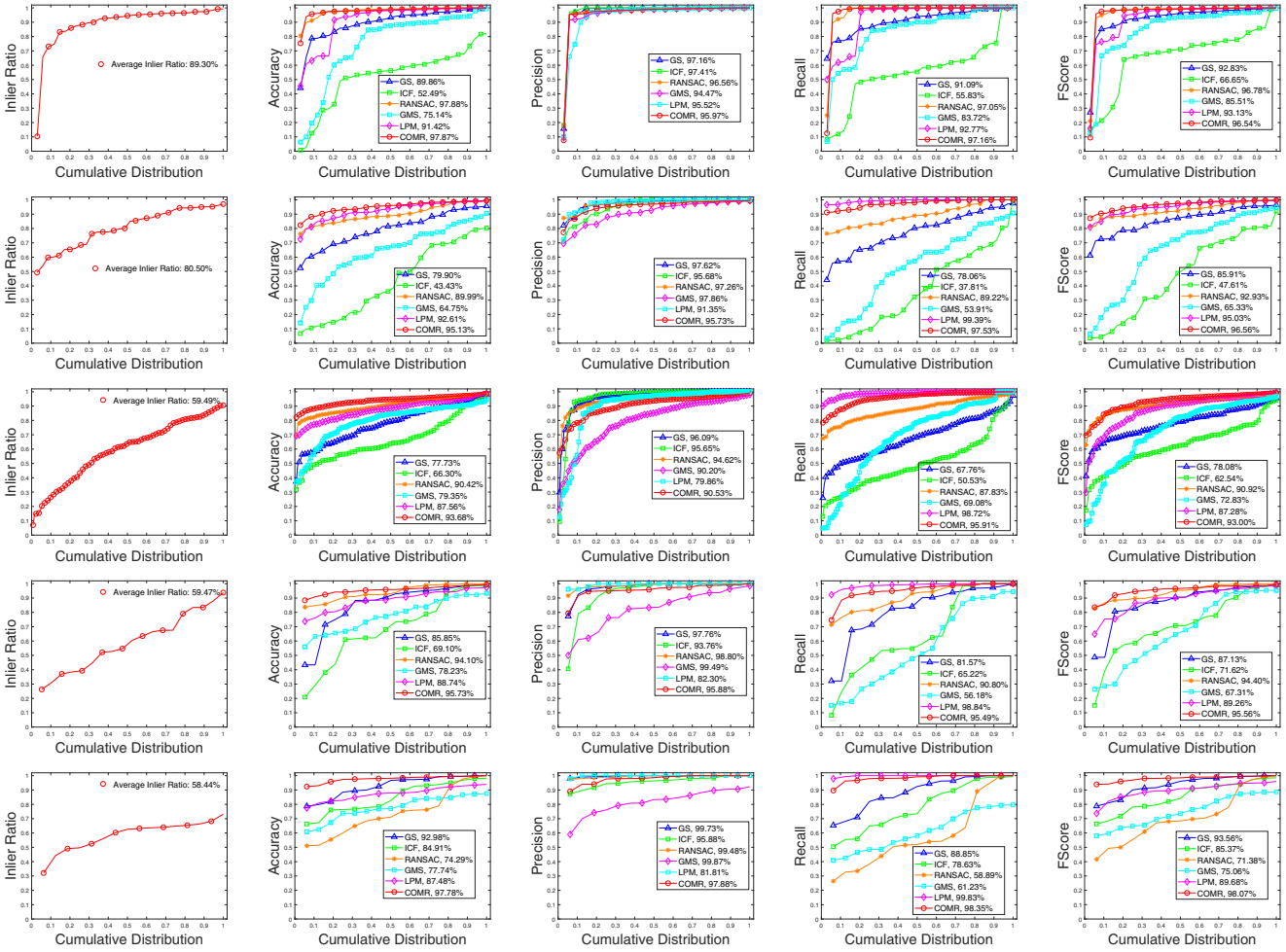

Fig. 6. Quantitative comparisons of RANSAC, GS, ICF, GMS, LPM and COMR on five datasets. From top to bottom: VGG, DAISY, DTU, AdelaideH and AdelaideF. From left to right: Initial inlier ratio, accuracy, precision, recall, and F-score with respect to the cumulative distribution.

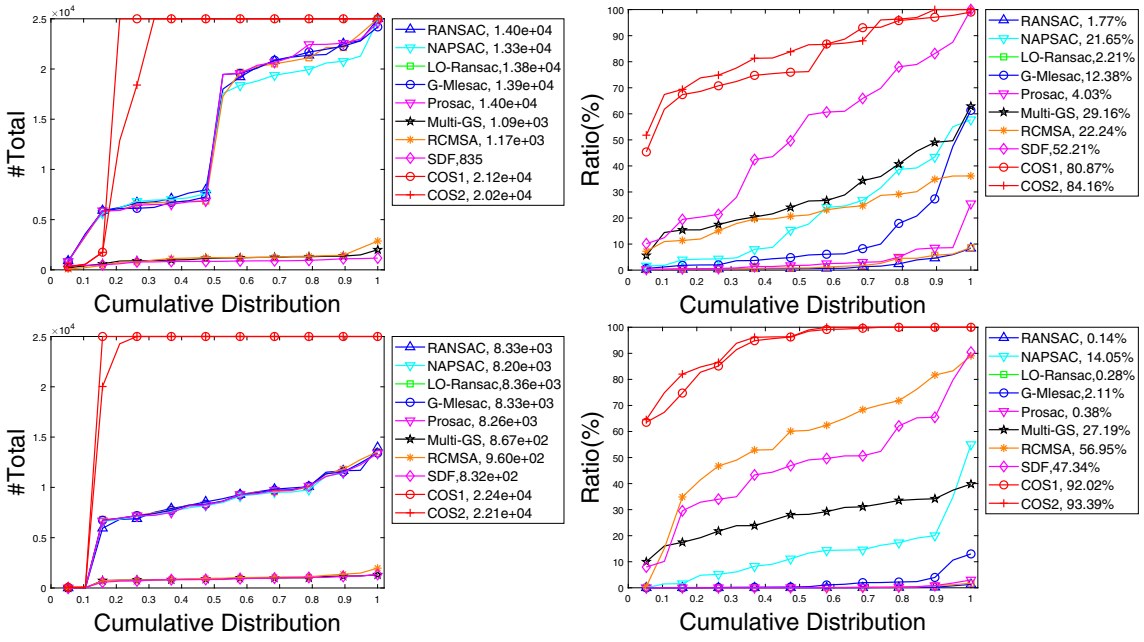

Fig. 7. Quantitative comparisons of ten sampling methods on all image pairs from the AdelaideH and AdelaideF dataset for homography based segmentation (Top) and two-view based motion segmentation (Bottom), respectively. (Left to Right) The total number of sampled minimal subsets and the value of ϱ with respect to the cumulative distribution.

ods, and they also can cover all model instances in data, especially for the image pairs with many model instances, e.g., "Unihouse" and "Cubebreadtoychips". This can valuate the effectiveness of the consensus of neighborhood elements and neighborhood topology proposed by our method. It is worth pointing out that, COS1/COS2 apparently conduct more operations than random sampling while they can sample more subsets than those random methods (e.g., RANSAC). The reason behind this is that, most of subsets sampled by RANSAC are degenerated, and cannot be used to generate a model hypothesis.

For COS1/COS2, we can see that, COS2 can further improve the value of ϱ over COS1, especially for homography estimation. This is because that, if two correspondences belong to a same model instance, then they will share the same length and angle, by which COS2 refines the sampled minimal subsets.

6. Discussion and conclusion

In this paper, we propose to exploit the co-occurrence statistics, which collect global statistics and local spatial information, to construct reliable neighborhoods for each correspondence. Then, we preserve the consensus of neighborhood elements and neighborhood topology, to mine consistent image feature correspondences for the mismatch removal problem. Experiments show that our mismatch removal method can achieve comparable performance to state-of-the-arts.

It is worth pointing out that, the computation complexity of our mismatch removal method is mainly governed by the step of collecting global co-occurrence statistics. Thus, the computation time of our method is related to the size of the input images. For most image pairs tested in the experiment, our method can obtain the solution within 1-2 s, which can be valuated by the experiment in model fitting (note that our sampling method achieves a large number of the minimal subsets within 5 s).

Based on our mismatch removal method, we further exploit the consensus of neighborhood elements and neighborhood topology to address geometric model fitting problems. The experimental results show that, our sampling method can achieve high-quality minimal subsets, which consist of a large proportion of all-inlier minimal subsets. Compared with several state-of-the-art sampling methods, our sampling method not only covers all model instances in data but also significantly improves the ratio of all-inlier minimal subsets to all sampled minimal subsets by several orders of magnitude.

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 by the National Natural Science Foundation of China under Grants 62072223, 61773295, and 61972187, by the Natural Science Foundation of Fujian Province under Grant 2020J01131199, and 2020J02024, and by the Fuzhou Science and Technology Project under Grant 2020-RC-186.

References

[1] J. Ma, X. Jiang, A. Fan, J. Jiang, J. Yan, Image matching from handcrafted to deep features: a survey, Int. J. Comput. Vis. 129 (1) (2021) 23–79.
[2] G. Xiao, J. Ma, S. Wang, C. Chen, Deterministic model fitting by local-neighbor preservation and global-residual optimization, IEEE Trans. Image Process. 29 (1) (2020) 8988–9001.
[3] J. Ma, X. Jiang, J. Jiang, Y. Gao, Feature-guided gaussian mixture model for image matching, Pattern Recognit. 92 (1) (2019) 231–245.
[4] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[5] J. Ma, J. Zhao, J. Jiang, H. Zhou, X. Guo, Locality preserving matching, Int. J. Comput. Vis. 127 (5) (2019) 512–531.
[6] C. Zhao, Z. Cao, C. Li, X. Li, J. Yang, NM-Net: mining reliable neighbors for robust feature correspondences, in: IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 215–224.
[7] J. Ma, X. Jiang, J. Jiang, J. Zhao, X. Guo, LMR: learning a two-class classifier for mismatch removal, IEEE Trans. Image Process. 28 (8) (2019) 4045–4059.
[8] R.J. Jevnisek, S. Avidan, Co-occurrence filter, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3184–3192.
[9] I. Shevlev, S. Avidan, Co-occurrence neural network, in: IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 4797–4804.
[10] R. Kat, R. Jevnisek, S. Avidan, Matching pixels using co-occurrence statistics, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1751–1759.
[11] H.S. Wong, T.-J. Chin, J. Yu, D. Suter, Dynamic and hierarchical multi-structure geometric model fitting, in: IEEE International Conference on Computer Vision, 2011, pp. 1044–1051.
[12] K. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, P. Fua, Learning to find good correspondences, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2666–2674.
[13] W. Sun, W. Jiang, E. Trulls, A. Tagliasacchi, K.M. Yi, ACNe: attentive context normalization for robust permutation-equivariant learning, in: IEEE Conference on Computer Vision and Pattern Recognition, 2020, p. 1128611295.
[14] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.
[15] L. Magri, A. Fusiello, T-Linkage: a continuous relaxation of j-linkage for multi–model fitting, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3954–3961.
[16] G. Xiao, H. Wang, Y. Yan, D. Suter, Superpixel-guided two-view deterministic geometric model fitting, Int. J. Comput. Vis. 127 (4) (2019) 323–329.
[17] H. Wang, G. Xiao, Y. Yan, D. Suter, Searching for representative modes on hypergraphs for robust geometric model fitting, IEEE Trans. Pattern Anal. Mach.Intell. 41 (3) (2019) 687–711.
[18] M. Pilu, A direct method for stereo correspondence based on singular value decomposition, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 1997, pp. 261–266.
[19] M. Pilu, A. Lorusso, Uncalibrated stereo correspondence by singular value decomposition, in: British Machine Vision Conference, IEEE, 1997, pp. 1–12.
[20] H. Liu, S. Yan, Common visual pattern discovery via spatially coherent correspondences, in: IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 1609–1616.
[21] X. Li, Z. Hu, Rejecting mismatches by correspondence function, Int. J. Comput. Vis. 89 (1) (2010) 1–17.
[22] W.Y. Lin, F. Wang, M.M. Cheng, S.K. Yeung, P.H.S. Torr, M.N. Do, J. Lu, CODE: coherence based decision boundaries for feature correspondence, IEEE Trans. Pattern Anal. Mach.Intell. 40 (1) (2018) 34–47.
[23] J. Bian, W. Lin, Y. Matsushita, S. Yeung, T.D. Nguyen, M. Cheng, GMS: grid-based motion statistics for fast, ultra-robust feature correspondence, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2828–2837.
[24] J. Bian, W. Lin, Y. Liu, L. Zhang, S. Yeung, M. Cheng, I. Reid, GMS: grid-based motion statistics for fast, ultra-robust feature, Int. J. Comput. Vis. 128 (2020) 1580–1593.
[25] C.R. Qi, O. Litany, K. He, L.J. Guibas, Deep hough voting for 3D object detection in point clouds, in: IEEE International Conference on Computer Vision, 2019, pp. 9277–9286.
[26] Z. Huang, X. Wang, L. Huang, C. Huang, W. Liu, CCNet: criss-cross attention for semantic segmentation, in: IEEE International Conference on Computer Vision, 2019, pp. 603–610.
[27] S. Bai, P. Tang, P.H. Torr, L.J. Latecki, Re-ranking via metric fusion for object retrieval and person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 740–749.
[28] W. Hu, Y. Huang, F. Zhang, R. Li, Noise-tolerant paradigm for training face recognition CNNs, in: IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 11887–11896.
[29] Yefeng Zheng, D. Doermann, Robust point matching for nonrigid shapes by preserving local neighborhood structures, IEEE Trans. Pattern Anal. Mach.Intell. 28 (4) (2006) 643–649.
[30] G. Xiao, H. Wang, T. Lai, D. Suter, Hypergraph modelling for geometric model fitting, Pattern Recognit. 60 (1) (2016) 748–760.
[31] T. Chin, J. Yu, D. Suter, Accelerated hypothesis generation for multistructure data via preference analysis, IEEE Trans. Pattern Anal. Mach.Intell. 34 (4) (2012) 625–638.
[32] O. Chum, J. Matas, Matching with PROSAC-progressive sample consensus, in: IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 220–226.
[33] O. Chum, J. Matas, J. Kittler, Locally optimized RANSAC, in: Joint Pattern Recognition Symposium, 2003, pp. 236–243.
[34] D. Nasuto, J.B.R. Craddock, NAPSAC: high noise, high dimensional robust estimation, in: British Machine Vision Conference, 2002, pp. 458–467.
[35] H. Le, T. Chin, A. Eriksson, T. Do, D. Suter, Deterministic approximate methods for maximum consensus robust fitting, IEEE Trans. Pattern Anal. Mach.Intell. 1 (1) (2020) 1–14.

[36] B.J. Tordoff, D.W. Murray, Guided-MLESAC: faster image transform estimation by using matching priors, IEEE Trans. Pattern Anal. Mach.Intell. 27 (10) (2005) 1523–1535.
[37] T. Pham, T. Chin, J. Yu, D. Suter, The random cluster model for robust geometric fitting, IEEE Trans. Pattern Anal. Mach.Intell. 36 (8) (2014) 1658–1671.
[38] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L.V. Gool, A comparison of affine region detectors, Int. J. Comput. Vis. 65 (1) (2005) 43–72.
[39] E. Tola, V. Lepetit, P. Fua, DAISY: an efficient dense descriptor applied to wide baseline stereo, IEEE Trans. Pattern Anal. Mach.Intell. 32 (5) (2010) 815–830.
[40] H. Aanes, R.R. Jensen, G. Vogiatzis, E. Tola, A.B. Dahl, Large-scale data for multiple-view stereopsis, Int. J. Comput. Vis. 120 (2) (2016) 153–168.

Guobao Xiao received the B.S. degree in information and computing science from Fujian Normal University, China, in 2013 and the Ph.D. degree in Computer Science and Technology from Xiamen University, China, in 2016. From 2016-2018, he was a Postdoctoral Fellow in the School of Aerospace Engineering at Xiamen University, China. He is currently a Professor at Minjiang University, China. He has published over 40 papers in the international journals and conferences including IEEE TPAMI/TIP/TITS/TIE, IJCV, PR, ICCV, ECCV, AAAI, etc. His research interests include machine learning, computer vision and pattern recognition. He has been awarded the best PhD thesis in Fujian Province and the best PhD thesis award in China Society of Image and Graphics (a total of ten winners in China). He also served on the program committee (PC) of CVPR, ICCV, ECCV, AAAI, WACV, etc. He was the General Chair for IEEE BDCLOUD 2019.

Shiping Wang received the Ph.D. degree from the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2014. He was a Research Fellow at Nanyang Technological University, Singapore, from 2015 to 2016. He is currently a Full Professor and Qishan Scholar with the College of Mathematics and Computer Science, Fuzhou University, Fuzhou, China. His research interests include machine learning, computer vision, and granular computing.

Wan Han is associate professor in Nanyang Technological University. He received his BEng from Northeastern Heavy Machinery Institute (Yanshan University) in China, and PhD from Leeds University in UK. He was teacher in Shanghai Institute of Technology, Visiting Scientist in Carnegie-Melon University, Research Officer in Oxford University after PhD. He joined NTI/NTU since 1992 as Lecture, Senior Lecture, and associate Professor. He has published 300 papers in international journals and conferences. He served and serves as associate editors for a few international journals, is now serving as Chair of Singapore chapter of Robotics and Automation Society, IEEE.

Jiayi Ma received the B.S. degree in information and computing science and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. From 2012 to 2013, he was an Exchange Student with the Department of Statistics, University of California at Los Angeles, CA, USA. He is currently a Professor with the Electronic Information School, Wuhan University. He has authored or co-authored more than 150 refereed journal and conference papers, including IEEE TPAMI/TIP, IJCV, CVPR, ICCV, ECCV, etc. His research interests include computer vision, machine learning, and pattern recognition. Dr. Ma has been identified in the 2020 and 2019 Highly Cited Researchers lists from the Web of Science Group. He is an Area Editor of Information Fusion and an Associate Editor of Neurocomputing.