

Deterministic Model Fitting by Local-Neighbor Preservation and Global-Residual Optimization

Guobao Xiao^{ID}, Jiayi Ma^{ID}, Member, IEEE, Shiping Wang^{ID}, Member, IEEE ,
and Changwen Chen^{ID}, Fellow, IEEE

Abstract—Geometric model fitting has been widely used in many computer vision tasks. However, it remains as a challenging task when handling multiple-structural data contaminated by noises and outliers. Most previous work on model fitting cannot guarantee the consistency of their solutions due to their randomness, precluding them from many real-world applications. In this research, we propose a fast two-view approximately deterministic model fitting scheme (called LGF), to provide consistent solutions for multiple-structural data. The proposed LGF scheme starts from defining preference function by preserving local neighborhood relationship, and then adopts the min-hash technique to roughly sample subsets. By this way, it is able to cover all model instances in data in the parameter space with a high probability. After that, LGF refines the previous sampled subsets by global-residual optimization. Furthermore, we propose a simple yet effective model selection framework to estimate the number and the parameters of model instances in data. Extensive experiments on real images show that the proposed LGF scheme is able to observe superior or very competitive performance on both accuracy and speed over several state-of-the-art model fitting methods.

Index Terms—Model fitting, local-neighbor preservation, global-residual optimization, min-hash, multiple-structure data.

I. INTRODUCTION

GEOMETRIC model fitting is a fundamental research topic in computer vision [1]–[6]. The goal of model fitting is to estimate the number and the parameters of model instances (also called “structures”) in data. As the data in real world often contain noises/outliers, how to effectively recover structures from the data with noises/outliers has become a

Manuscript received February 22, 2020; revised July 14, 2020 and August 23, 2020; accepted August 29, 2020. Date of publication September 17, 2020; date of current version September 23, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61702431, Grant 61773295, and Grant 61972187; in part by the Natural Science Foundation of Fujian Province under Grant 2020J01131199, and in part by the Fuzhou Science and Technology Project under Grant 2020-RC-186. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Chun-Shien Lu. (*Corresponding author: Jiayi Ma*)

Guobao Xiao is with the Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, College of Computer and Control Engineering, Minjiang University, Fuzhou 350108, China, and also with The State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China.

Jiayi Ma is with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: jyma2010@gmail.com).

Shiping Wang is with the College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China.

Changwen Chen is with the School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen 518172, China, and also with the Department of Computer Science and Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260-1660 USA.

Digital Object Identifier 10.1109/TIP.2020.3023576

challenging task. Moreover, it is common in many computer vision tasks (such as motion segmentation [4], homography/fundamental matrix estimation [3], and 3D reconstruction [7]) where the data often contain multiple structures, which inevitably interfere with each other during the model fitting process.

Sampling minimal subsets is one key step of model fitting, where a model hypothesis is generated for model selection. Here, a minimal subset contains the minimum number (p) of data points required to estimate a model hypothesis (e.g., 2 for line fitting and 4 for homography fitting). The minimal subset 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 an outlier (gross outlier or pseudo outlier¹). 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 (p). For example, RANSAC [8] is one of the most successful sampling algorithms due to its simplicity and effectiveness. However, to hit a model instance with the probability β , RANSAC need to sample $\frac{\log(1-\beta)}{\log(1-(1-\alpha)^p)}$ minimal subsets, where α denotes the outlier ratio. That is, RANSAC requires to sample a large number of minimal subsets when the data contain a large proportion of outliers. Therefore, randomly sampling becomes intractable in practice.

Over the years, many guided sampling algorithms [3], [9]–[14] have been proposed to reduce the number of sampled minimal subsets while hitting all structures with reasonable time. But these sampling algorithms still cannot guarantee the consistent solutions due to their random nature. To obtain consistent solutions, some deterministic sampling algorithms (e.g., [3], [14]–[20]) are proposed. These deterministic fitting methods are able to significantly improve the stability of fitting results over random sampling-based fitting methods. Nevertheless, it remains challenging to customize an effective and efficient deterministic fitting method for practical use. Note that, most existing deterministic fitting methods only work for single-structure data, and their computational efficiency is far from satisfactory.

In this research, we present a two-view “approximately” deterministic model fitting scheme, based on Local-neighbor

¹A pseudo outlier is an inlier belonging to the other model instance.

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

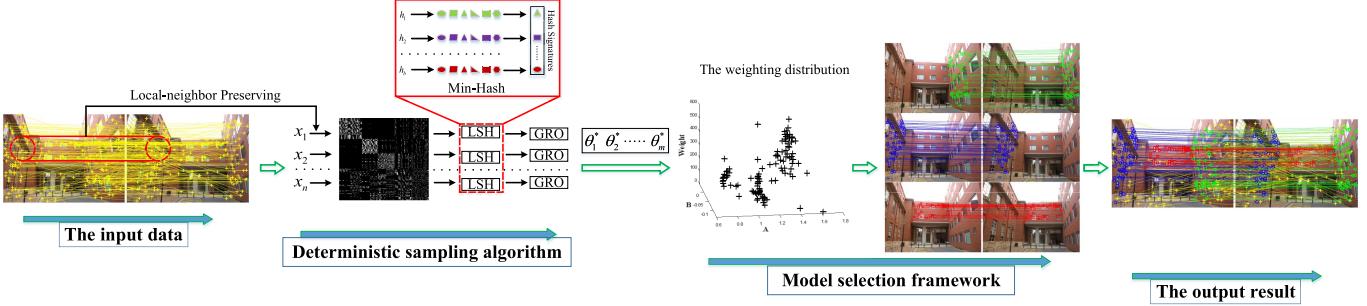


Fig. 1. Overview of the proposed scheme for homography-based segmentation.

Preservation and Global-residual Optimization, aiming at improving the effectiveness and efficiency of the fitting performance for multiple-structural data. We observe that effective correspondences (i.e., inliers) have similar local neighborhood relationships in two-view images; While ineffective correspondences (i.e., outliers) have very incoherent nearest neighbors, due to physical constraints. We then define preference function of input data points based on such observations. Subsequently, we adopt the preference analysis to promote the extraction of close correspondences by searching for correspondences with similar neighborhood. To further improve the efficiency, we adopt min-hash [21], a locality-sensitive hashing (LSH) [22] procedure, to sample minimum subsets, where each data point is selected to be a seed of a sampled minimum subset to hit all model instances in data in the parameter space. To refine the previous sampled minimum subsets (promoting spatial proximity usually has the drawback of causing degeneracy), we further introduce a global-residual optimization (GRO) strategy.

In addition, we also propose a simple yet effective model selection framework by adopting the effective sampled minimum subsets, to estimate the number and the parameters of model instances in data. We show in Fig. 1 an overview of the proposed Local-neighbor preservation and Global-residual optimization based Fitting (LGF) scheme as applied to homography-based segmentation.

We summarize the key contributions of the proposed scheme as follows:

- We propose to adopt preference analysis and the min-hash technique to exploit local-neighbor preservation relationships, for approximately deterministically sampling a small number of effective minimal subsets. To the best of our knowledge, we are the first one to gather neighborhood information of feature points but not model hypotheses to define the preference function for model fitting. As a result, we can directly initialize some good minimum subsets with stability; while most of existing guided sampling algorithms randomly initialize minimum subsets, that is, they cannot guarantee the quality and stability of minimum subsets.
- We propose a global-residual optimization strategy by analyzing the elements of the sampled minimal subsets to refine the sampled minimal subsets. To improve the effectiveness of GRO, we propose to consider the

consensus of the corresponding minimum subset to measure the quality of a model hypothesis.

- We propose a simple yet effective model selection framework for model fitting. The framework introduces the inlier noise scale constraint and the mutual information theory [23], to estimate the number and the parameters of model instances in data.

Extensive qualitative and quantitative experiments have been performed on real images. The results show that the proposed LGF scheme is able to obtain more accurate fitting results over several state-of-the-art model fitting methods with much less computation time.

It is worth pointing out that we can only call the proposed LGF scheme as an “approximately” deterministic scheme, since that LSH (used by the proposed LGF) is an approximate algorithm, and we cannot in theory guarantee the deterministic nature. However, in practice, LGF can provide very consistent solutions for the model fitting problems (see more details in Sec. V-A.2).

The rest of the paper is organized as follows: In Sec. II, we review the related literatures. In Sec. III, we propose an approximately deterministic sampling algorithm for model hypothesis generation. Then, based on the model hypotheses, we propose a novel model selection framework in Sec. IV. In Sec. V, we present the experimental results. We further discuss and analyze the proposed LGF scheme in Sec. VI, and draw conclusions in Sec. VII.

II. RELATED WORK

In this section, we briefly review the related model fitting methods for the current study according to the two steps of a general model fitting framework. This related work includes two method types: the first type focuses on the subset sampling, whereas the second type focuses on the model selection.

A. Sampling Algorithms

RANSAC, which randomly samples minimal subsets from input data points, is one of the most popular sampling algorithms. However, it often requires to sample a large number of minimal subsets to hit all structures when the data include a large proportion of outliers, and these sampled minimal subsets often include a large proportion of bad ones (whose elements

are not all inliers belonging to the same model instance). Many guided sampling algorithms, e.g., [2], [9]–[13], [24]), are proposed to sample high-quality minimal subsets for model hypothesis generation.

According to the guided way in the sampling process, these sampling algorithms can be roughly classified into two categories, i.e., the prior information-based and the posterior information-based algorithms. The prior information-based sampling algorithms, e.g., Proximity [12], NAPSAC [13] and PROSAC [10], require to provide the prior information (e.g., spatial distance, matching scores and superpixel information) in advance to guide the sampling process. While the posterior information-based sampling algorithms, e.g., MultiGS [9], LO-RANSAC [11], HMSS [24] and CBS [2], use the information derived from the minimal subsets sampled in previous process to guide the following process.

These guided sampling algorithms can provide high quality of minimal subsets for simple data, however, they cannot guarantee the consistency especially for multiple-structural data with a large proportion of outliers. To address this issue, some deterministic sampling algorithms (e.g., [3], [14]–[20], [25]) are proposed. Most of deterministic fitting methods, e.g., [15], [17]–[20], [25]), propose to solve a global optimization problem to guarantee the consistency. Nevertheless, they only work for single-structural data. [3], [14], [16] can deal with multiple-structural data, but they still have obvious limitations. For example, [16] requires to repeatedly generate a large number of model hypotheses, which may results in computational inefficiency; [3], [14] require to provide the superpixel information in advance, and thus they cannot work well for the data with complex scene due to the superpixel limitation.

The proposed sampling algorithm in this research samples the minimal subsets from coarse to fine, and it is one of deterministic sampling algorithms. However, the proposed sampling algorithm is significantly different with the previous sampling algorithms. Recall that the proposed sampling algorithm adopts preference analysis and the min-hash technique to exploit local-neighbor preservation relationships, to sample a small number of effective minimal subsets. That is, the proposed sampling algorithm is more general and it also can be implemented in parallel to further improve the sampling efficiency.

B. Model Selection Algorithms

The existing model selection algorithms can be roughly classified into two categories, i.e., the consensus analysis-based and the preference analysis-based algorithms, according to the selecting way in the model selection process. The preference analysis-based model selection algorithms, e.g., SWS [26], HF [27], RPA [28], J-linkage [29], T-linkage [30] and MCT [1], estimate model instances by analyzing the relationship between data points. For example, SWS introduces large hyperedges to obtain the preference information of data points. HF adopts a no-uniform hypergraph to represent the relationship between model hypotheses and data points. RPA proposes to adopt principal component analysis and non-negative matrix factorization to exploit the relationship

between data points. J-linkage/T-linkage use different distance measures to cluster data points based on the preference information of data points. MCT extends T-linkage to handle different nested classes of models. These model selection algorithms are able to generally achieve good performance on the fitting accuracy but they often suffer from high computational complexity.

The consensus analysis-based model selection algorithms, e.g., RansaCov [31], RHT [32], GPbM [33], AKSWH [34] and MSHF [35], directly select model instances from the generated model hypotheses in the parameter space. This type of model selection algorithms selects model instances according to different evaluation criterions, e.g., the number of estimated inliers and the weighting scores. For example, RansaCov selects model instances that include the maximum number of inliers. RHT, AKSWH and MSHF select model instances by defining different weighting functions. GPbM selects model instances according to an optimization function. There are also some energy-based model fitting methods, e.g., RCMSA [36], EMSAC [37], CORAL [38] and Prog-X [39]. These model selection algorithms are able to obtain good fitting accuracy if we input high-quality model hypotheses.

In this research, we propose a consensus analysis-based model selection algorithm, which selects model instances by analyzing the weighting scores of model hypotheses and the relationship between the model instances estimated in the previous process and the remaining model hypotheses. Compared with the existing model selection algorithms, the proposed algorithm is more general and much simpler to be implemented. Furthermore, the proposed algorithm is able to estimate the number of model instances in data (note that some of existing algorithms, e.g., RansaCov and RHT, require users to provide the number of model instances in advance).

III. PROPOSED APPROXIMATELY DETERMINISTIC SAMPLING ALGORITHM

In this section, we describe the details of the proposed approximately deterministic sampling algorithm. To this end, we first search K nearest neighbors for each feature point in two-view images, respectively. Then, based on the neighborhood relationships, we construct binary preference sets to represent data points (i.e., correspondences). After that, we select each data point as a seed of a minimal subset, and then, based on the preference sets, we adopt minhash to sample other elements of each minimal subset. At last, we refine the minimal subsets by global-residual optimization (GRO).

A. Local-Neighbor Preservation for Sampling Minimal Subsets

In real-world data, the local neighborhood relationships among feature points in a small region will be preserved better than the absolute distance due to physical constraints, especially for viewpoint changing [40]. Thus, for inliers, they will share similar neighbors in two views; In contrast, for outliers, their neighbors will be significantly different in

two views. Based on this observation, we construct a binary preference set to represent each data point.

Specifically, for a data point $x_i = (\hat{x}_i, \hat{y}_i) \in X$ in two-view images, we search K nearest neighbors $\mathcal{N}_{\hat{x}_i}$ and $\mathcal{N}_{\hat{y}_i}$ of the feature points \hat{x}_i and \hat{y}_i , respectively. Then, given n data points, we construct an $n \times 2n$ binary matrix P by defining its (i, j) -th entry in one view as:

$$P(i, j) = \begin{cases} 1, & \text{if } x_j \in \mathcal{N}_{\hat{x}_i}, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

and $P(i, 2 * j)$ is accordingly defined based on $\mathcal{N}_{\hat{y}_i}$ in the other view.

Then, each data point x_i can be represented by a row P_i , i.e., the preference set $PS(x_i)$. The main intuition is that, if two data points are inliers belonging to a same model instance, they will have similar neighbors, i.e., preference sets. Thus, we can use the preference information to sample a minimal subset. Here, to hit all model instances in data with a high probability, we firstly select each data point as a seed of a minimal subset, and then search the $p - 1$ data points with the most similar preference sets as the seed to be the other elements of the minimal subset. The similarity between two data points (x_i and x_j) can be measured by the Jaccard distance [29]:

$$D(x_i, x_j) = \frac{|PS(x_i) \cup PS(x_j)| - |PS(x_i) \cap PS(x_j)|}{|PS(x_i) \cup PS(x_j)|}. \quad (2)$$

To efficiently compute the Jaccard distance between two data points, we approximate it as computing the Hamming distance of the respective hash signature.³ Note that locality-sensitive hashing (LSH) [22] is a powerful tool for approximate nearest-neighbor search, and it is able to efficiently map the similar ones of the input items to the same bucket of a hash table with a high probability, while map the others to different buckets. Here, we use LSH to map the similar data points (which are the inliers of the same model hypothesis with a high probability) to the same bucket (i.e. the same minimal subset for model fitting).

Specially, following [41], [42], we define the LSH family as: Given the input data points $X = \{x_1, x_2, \dots, x_n\}$ and the corresponding distance measure D between data points, an LSH is a probability distribution on family H of hash functions h , and each h satisfies the property:

$$Pr(h(x_i) = h(x_j)) = D(x_i, x_j), \quad (3)$$

where $h(x) \in \{h_i(x)\}_{i=1}^b$ is a bijective random function. Considering preference sets used to compute the Jaccard distance are binary sets, we use min-hash [21], i.e., one of LSH, to compute the hash signature.

The intuition of min-hash is described as follows: each data point is mapped to a hash key constructed by the b hash functions, and if two data points are similar enough, they will be assigned to the same hash key. The number of hash functions has been discussed in [43], and the hash functions

³A hash signature is collected by b hash values, which is computed by the b hash functions.

can be implemented as a linear transformation on a convenient finite field [43], which can obtain consistent results.

In our formulation, we select $p - 1$ data points with the most similar preference sets as the seed to be the other elements of the minimal subset. However, to improve the quality of minimal subsets, we propose a filtering strategy to remove some impure minimal subsets. Specifically, we first compute the initial elements of a minimal subset for each seed x_i under Jaccard distance:

$$\mathcal{MS}_{x_i} = \{x_j | D(x_i, x_j) < \xi, x_j \in X\}, \quad (4)$$

where ξ is a nonzero parameter. Then select $p - 1$ data points from the initial elements of each minimal subset. It is worth pointing out that, the filtering strategy is able to improve the quality of the sampled minimal subsets. This is because, if a seed is an inlier, we can obtain enough elements to sample a minimal subset for model hypothesis generation; In contrast, if a seed is an outlier, it is hard to obtain enough elements.

Note that, [43] also uses min-hash to deal with the model fitting problem as the proposed scheme. However, the proposed scheme uses min-hash to sample minimal subsets while [43] uses min-hash to cluster data points for model selection. Meanwhile, the proposed scheme can be implemented in parallel while [43] requires to make a large number of iterations. In addition, the proposed scheme adopts a filtering strategy to remove some impure minimal subsets, and this will be more appropriate for addressing the general model fitting problem. Therefore, the proposed scheme is significantly different with [43].

B. Global-Residual Optimization for Refining Minimal Subsets

In the last subsection, we use local-neighbor preservation information to sample minimal subsets for model fitting, which covers all model instances in data with a high probability. To further improve the quality of sampled minimal subsets, we adopt the global-residual optimization (GRO) strategy, used in [2], [3], [24], by exploiting the residual values between model hypotheses (generated by the sampled minimal subsets) and data points.

The original GRO strategy is proposed in [24]. The main steps of GRO strategy are described as follows: For a sampled subset, sort the residual values between a model hypothesis (generated by the sampled subset) and the input data points in ascending order. Then, sample the data points which are around the m_k -th point in the order of residual values. Here m_k is the minimum cluster size (provided in advance). Repeated these two steps until it converges to a solution.

The original GRO strategy is able to effectively refine sampled minimal subsets within a few iterations. However, it may always converge to a model instance for the data with multiple model instances since the input minimal subsets are randomly sampled. To deal with this problem, [2] proposes a modified version of the iterative procedure. In particular, it samples the subset from the data points which are not inliers of the model hypothesis converged in the previous procedure. To improve the fitting performance, [3] introduces a weighting

score to measure the quality of the model hypothesis, and selects the model hypothesis with the maximum score instead of the converged one.

Although all of [2], [3], [24] have refined the sampled minimal subsets to some extent, they have some obvious limitations. For example, all of them require to provide the number of model instances in advance, and the performance of [3] depends on the accuracy of the weighting measure.

In this research, we improve the original GRO strategy by analyzing the sampled minimal subsets to improve the accuracy of the weighting measure. Specifically, for a model hypothesis θ_i generated by a sampled subset $\widetilde{\mathcal{MS}}_{x_i}$, we define its weighting score as follows:

$$w(\theta_i) = \underbrace{\frac{1}{n} \sum_{j=1}^n \frac{\mathbf{EK}(r_i^j/bw(\theta_i))}{\tilde{s}(\theta_i)bw(\theta_i)}}_{(a)} \underbrace{Sd(\widetilde{\mathcal{MS}}_{x_i})}_{(b)}, \quad (5)$$

where $\tilde{s}(\theta_i)$ is the estimated inlier noise scale of the i -th model hypothesis; r_i^j is the residual derived from the i -th model hypothesis and the j -th keypoint correspondence; $bw(\theta_i)$ is the bandwidth of the i -th model hypothesis defined as [34]

$$bw(\theta_i) = \left[\frac{243 \int_{-1}^1 \mathbf{EK}(\lambda)^2 d\lambda}{35n \int_{-1}^1 \lambda^2 \mathbf{EK}(\lambda) d\lambda} \right]^{0.2} \tilde{s}(\theta_i); \quad (6)$$

For the kernel function $\mathbf{EK}(\cdot)$, we employ the popular Epanechnikov kernel, which is written as follows:

$$\mathbf{EK}(\lambda) = \begin{cases} 0.75(1 - \|\lambda\|^2), & \|\lambda\| \leq 1, \\ 0, & \|\lambda\| > 1. \end{cases} \quad (7)$$

$Sd(\widetilde{\mathcal{MS}}_{x_i})$ is a penalty term based on the consensus of the minimal subset $\widetilde{\mathcal{MS}}_{x_i}$. We define the consensus of a minimal subset $\widetilde{\mathcal{MS}}_{x_i}$ based on the ration of length and the angle between an element x_j and $\widetilde{\mathcal{MS}}_{x_i}$:

$$Cs(x_j, \widetilde{\mathcal{MS}}_{x_i}) = \frac{\min\{|x_j|, |med_{x_i}|\}}{\max\{|x_j|, |med_{x_i}|\}} \cdot \frac{(x_j, med_{x_i})}{|x_j| \cdot |med_{x_i}|}, \quad (8)$$

where med_{x_i} is the median value of all elements in $\widetilde{\mathcal{MS}}_{x_i}$, and (\cdot, \cdot) denotes the inner product. The reason behind Eq. (8) is that, if two data points are inliers belonging to a same model instance, they will have similar lengths and directions [44]. Then, a larger value of $Cs(x_j, \widetilde{\mathcal{MS}}_{x_i}) \in [-1, 1]$ denotes higher consensus of the minimal subset. Thus, $Sd(\widetilde{\mathcal{MS}}_{x_i})$ is computed by the standard deviation of the consensus, i.e., $Sd(\widetilde{\mathcal{MS}}_{x_i}) = std(\{1/Cs(x_j, \widetilde{\mathcal{MS}}_{x_i})\}_{x_j \in \widetilde{\mathcal{MS}}_{x_i}})$.

In Eq. (5), (a) is the term proposed in [34], and (b) is a penalty term based on the consensus of the elements in a minimal subset. For (a), θ_i will be assigned a higher weighting scores if it contains a larger number of inliers and with smaller residual values; For (b), θ_i will be punished (i.e., the score becomes smaller) if its sampled subset contains more outliers.

The penalty term in Eq. (5) is able to improve the accuracy of the weighting measure especially for the data that contain a model instance with a small number of inliers. This is because, a model hypothesis will be assigned a small score by (a) if

Algorithm 1 The Approximately Deterministic Sampling Algorithm

Input: Data points $X = \{x_1, x_2, \dots, x_n\}$, K for preference set generation, b hash functions, ξ for minimal subset selection, and minimum cluster size (m_k).

- 1: $l_{max} \leftarrow 50$, $h \leftarrow p + 2$.
- 2: Compute K nearest neighbors of each feature point of input data points.
- 3: Generate the preference sets of each data point by using Eq. (1).
- 4: Compute the hash signatures using the preference sets.
- 5: Obtain the Hamming distance between each two data points.
- 6: **for** $i = 1$ to n **do**
- 7: Compute the initial elements \mathcal{MS}_{x_i} by using Eq. (4).
- 8: **if** $|\mathcal{MS}_{x_i}| < p - 1$ **then continue; end if**
- 9: Generate a hypothesis θ_i^0 using x_i and \mathcal{MS}_{x_i} .
- 10: $l \leftarrow 1$.
- 11: **repeat**
- 12: $[rank_j^l]_{j=1}^n \leftarrow SortRes(X, \hat{\theta}_i^l)$;
- 13: Compute the weighting score by using Eq. (5);
- 14: Evaluate the stoping criterion (l_{stop}) [3];
- 15: **if** l_{stop} **then break; end if**
- 16: $X_{s_i}^{l+1} \leftarrow [x_{rank_j^l}]_{j=m_k-h+1}^{m_k}$ // Sample a new subset.
- 17: $\hat{\theta}_i^{l+1} \leftarrow LSFit(X_{s_i}^{l+1})$ // Fit a new hypothesis.
- 18: **until** $(l + + > l_{max})$
- 19: $\theta_i^* \leftarrow argmax\{w(\theta_i^l)\}_{l=1,2,\dots}$
- 20: **end for**

Output: The model hypothesis set θ^* ($=\{\theta_i^*\}_{i=1}^m$)

it does not contain a large number of inliers, even when all elements of its sampled subset are inliers; While the penalty term will preserve its score for this case.

With all the above ingredients, we summarize the approximately deterministic sampling algorithm in Algorithm 1.

IV. PROPOSED MODEL SELECTION FRAMEWORK

In this section, we propose a novel model selection framework to estimate the number and the parameters of model instances in data.

One of the most popular model selection framework is the “fit-and-remove” framework, which can be described as: Sequentially sample model hypotheses, select one model instance, and remove the inliers of the selected model instances from the input data points. However, the framework requires to repeatedly generate model hypotheses, which is very time-consuming. To deal with this problem, an improved “fit-and-remove” framework is proposed [3]: Sample model hypotheses, sequentially select one model instance, and remove redundant model hypotheses from the generated model hypotheses according to the selected model instances.

We can see that, the improved version is able to avoid repeatedly generating model hypotheses. However, it still requires to provide the number of model instances in advance to terminate the iterations. Note that the estimation of the

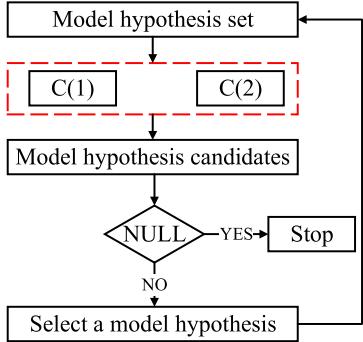


Fig. 2. The framework of the proposed model selection.

Algorithm 2 The Proposed Model Selection Framework

Input: Data points $X = \{x_1, x_2, \dots, x_n\}$, the model hypothesis set θ^* ($=\{\theta_i^*\}_{i=1}^m$) and minimum cluster size (m_k).
1: Initialize model hypothesis candidates $\theta^{**} \leftarrow \theta^*$.
2: **repeat**
3: Select a model hypothesis from θ^{**} as the estimated model instance $\tilde{\theta}$.
4: Compute the first condition by using Eq. (9).
5: Compute the second condition by using Eq. (10).
6: Update the model hypothesis candidates θ^{**} according to the two conditions.
7: **until** ($\theta^{**} = \text{NULL}$)
Output: The estimated model instance $\tilde{\theta}$ ($=\{\tilde{\theta}_i\}_{i=1}^m$)

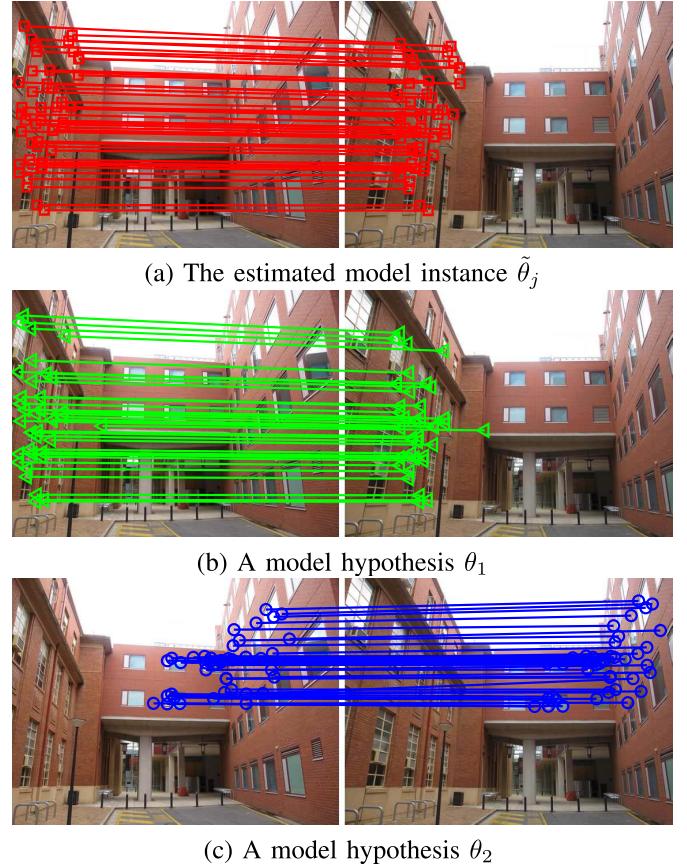
number of model instances in data is also an important task for the model fitting problem.

To relieve the above limitations, we propose a novel model selection framework based on the inlier scale estimation and the mutual information theory [23]. We find that if a model hypothesis is the true model instance in data, it should satisfy two conditions, i.e., enough inliers and significant differences from the other estimated model instances on the inlier distributions centered. Thus, we use the inlier scale estimation to judge if the model hypothesis has enough inliers (i.e., minimum cluster size m_k), and the mutual information theory to judge if two model hypotheses correspond to the same model instance.

Specifically, as shown in Fig. 2, for the input model hypothesis set, we filter candidates by using two conditions, i.e., enough inliers ($C(1)$ in Fig. 2) and significant differences from the other estimated model instances ($C(2)$ in Fig. 2). Then we select the model hypothesis with the maximum weighting score from the candidates as the estimated model instance if the candidates are not null; Otherwise, we terminate the process.

For $C(1)$, we adopt the inlier noise scale estimators, e.g., ALKS [45], MSSE [46], TSSE [47] and IKOSE [34], to estimate the inliers. Here, we use IKOSE to do this due to its efficiency and simplicity of implementation. Thus, for a model hypothesis θ_i , $C_{\theta_i}(1)$ is written as follows:

$$C_{\theta_i}(1) = \begin{cases} 1, & \text{if } |\{r_i^j\}_{j=1}^n < 2.5\tilde{s}(\theta_i)\} | > m_k, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

Fig. 3. An example of mutual information for model fitting. (a)-(c) Three model hypotheses θ_j , θ_1 and θ_2 . (d) The conditional probability of data points with model hypotheses and the mutual information between model hypotheses.

By Eq. (9), the model hypothesis candidates with enough inliers ($> m_k$) will be kept; While the others will be removed from the candidates.

For a model hypothesis θ_i , the condition $C_{\theta_i}(2)$ is written as follows:

$$C_{\theta_i}(2) = \begin{cases} 1, & \text{if } \bigcap_{j=1,2,\dots} MI(\theta_i, \tilde{\theta}_j) < 0, \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

where $\tilde{\theta}_j$ is the model instance estimated in the previous iterations. $MI(\theta_i, \tilde{\theta}_j)$ is the mutual information between two

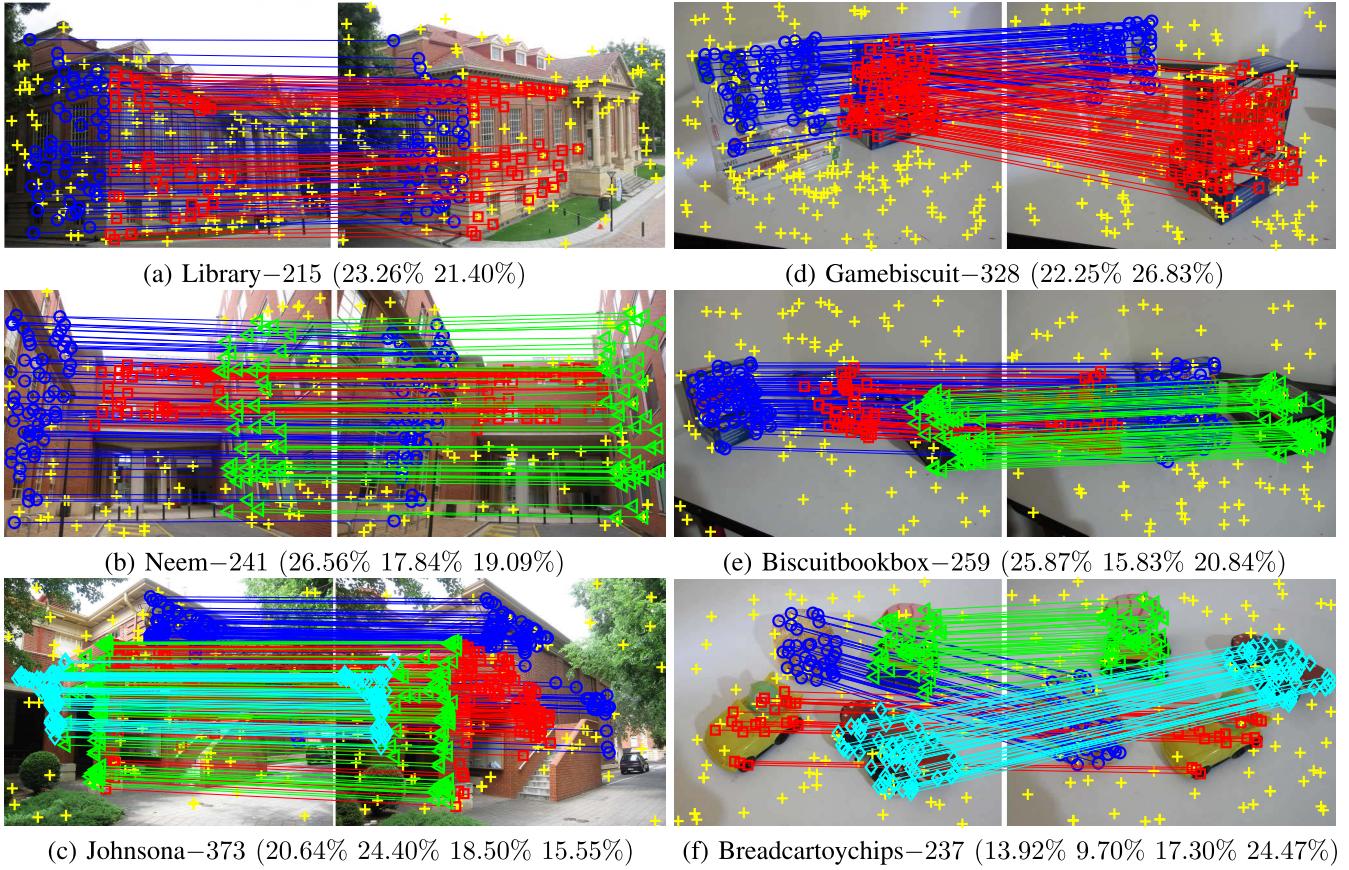


Fig. 4. Six representative image pairs ((a)-(c) for homography-based segmentation and (d)-(f) for two-view-based motion segmentation) from the AdelaideRMF dataset. We show the total number of correspondences (and the inlier ratios for each model instance).

model hypotheses θ_i and $\tilde{\theta}_j$, and it is defined as:

$$MI(\theta_i, \tilde{\theta}_j) = \log \frac{n \sum_{q=1}^n \rho(x_q|\theta_i) \rho(x_q|\tilde{\theta}_j)}{\sum_{q=1}^n \rho(x_q|\theta_i) \sum_{q=1}^n \rho(x_q|\tilde{\theta}_j)}, \quad (11)$$

where $\rho(x_q|\theta_i)$ is defined as:

$$\rho(x_q|\theta_i) = \frac{1}{\tilde{s}(\theta_i)\sqrt{2\pi}} \exp\left(-\frac{(r_i^q)^2}{2\tilde{s}(\theta_i)^2}\right). \quad (12)$$

Eq. (12) is based on the assumption of a Gaussian inlier noise model [34], and it represents the conditional probability of a data point x_q with a model hypothesis θ .

Then, in Eq. (11), we compute the mutual information between two model hypotheses θ_i and $\tilde{\theta}_j$ by measuring the similarity between the inlier distributions centered on θ_i and $\tilde{\theta}_j$. Specifically, if $MI(\theta_i, \tilde{\theta}_j)$ is larger than zero, then θ_i and $\tilde{\theta}_j$ are statistically dependent; Otherwise, they are statistically independent. That is, if two model hypotheses correspond to the same model instance (their inlier distributions are similar), they will share large mutual information; Otherwise, they will share small mutual information.

To illustrate the mutual information, in Fig. 3, we show an example on the “Neem” image pair. From Fig. 3(d), we can see that, Eq. (12) is able to capture the inlier distributions of model hypotheses. For the estimated model instance $\tilde{\theta}_j$ (Fig. 3(a)) and two model hypotheses θ_1 and θ_2 (Fig. 3(b) and 3(c)), $\tilde{\theta}_j$ and θ_1 share similar inlier distributions, then

$MI(\theta_1, \tilde{\theta}_j) = 1.55$; While $\tilde{\theta}_j$ and θ_2 share different inlier distributions, then $MI(\theta_2, \tilde{\theta}_j) = -60.18$.

Thus, by Eq. (10), the model hypothesis candidates corresponding to the same model instance with the model instances estimated in the previous iterations will be removed; While the others that have significant differences from the model instances estimated in the previous iterations will be kept.

With all the above ingredients, we summarize the proposed model selection framework in Algorithm 2. Compared with the traditional “fit-and-remove” framework and the improved version [3], the proposed framework not only does not require to repeatedly generate model hypotheses, but also is able to automatically estimate the number of model instances in data. What is more, the proposed framework introduces the mutual information theory, which is able to show the similarity of the inlier distributions centered on two model hypotheses, to judge if a model hypothesis is a candidate of the estimated model instance. In contrast, [3] adopts the information of sampled minimal subsets to remove model hypotheses, and thus the performance of [3] largely depends on the quality of the sampled minimal subsets.

V. EXPERIMENTS

In this section, we investigate the performance of the proposed LGF fitting scheme on real images for the subset sampling and model selection tasks. All experiments are run on MS Windows 10 with Intel Core i7-8565 CPU 1.8GHz

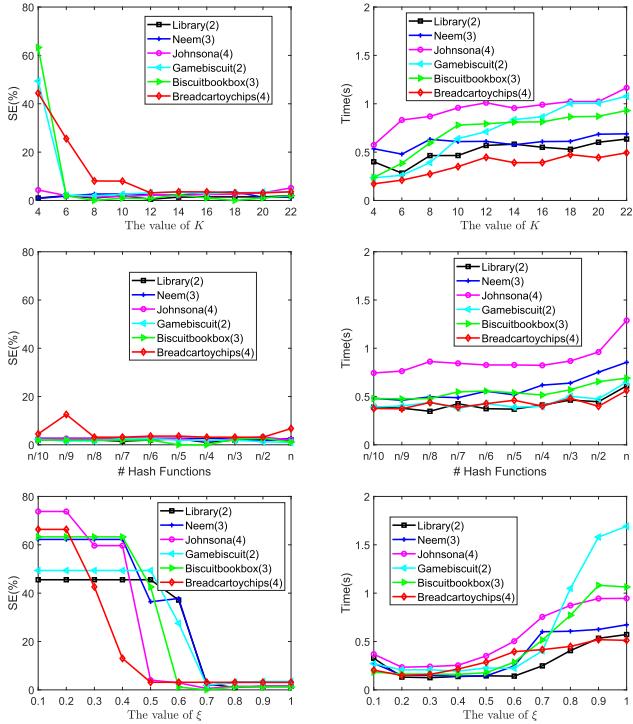


Fig. 5. The segmentation errors and the CPU time obtained by the proposed LGF fitting scheme with different parameter values on six representative image pairs.

and 16GB RAM. The segmentation error (SE) is computed as follows [31], [33]:

$$SE = \frac{\# \text{mislabeled data points}}{\# \text{data points}} \times 100\%. \quad (13)$$

A. Algorithm Analysis

In this subsection, we analyze the parameters, the consistency and some main components of the proposed LGF schemes.

1) *Parameter Analysis and Settings:* There are four parameters for LGF, i.e., K for preference set generation, the number of hash functions b , the minimal subset selection threshold ξ , and the minimum cluster size (m_k). We test different values of the first three parameters for homography-based segmentation and two-view-based motion segmentation. For each task we respectively test three image pairs, which include different numbers of model instances (see Fig. 4), from the Adelai-deRMF dataset [48]. This dataset contains 19 image pairs for homography-based segmentation, and 19 image pairs for two-view-based motion segmentation, and the correspondences are generated by SIFT matching [49]. Then we show the segmentation errors and the CPU time obtained by LGF with different parameter values in Fig. 5.

We can see that, when we set a small value of K , LGF obtains large segmentation errors on the three image pairs for two-view-based motion segmentation; While the segmentation errors on the three image pairs for homography-based segmentation have not significant changes. This is because, sampling a minimal subset for two-view-based motion segmentation requires 7 or 8 correspondences at least and for homography-based segmentation only requires 4 correspondences, then

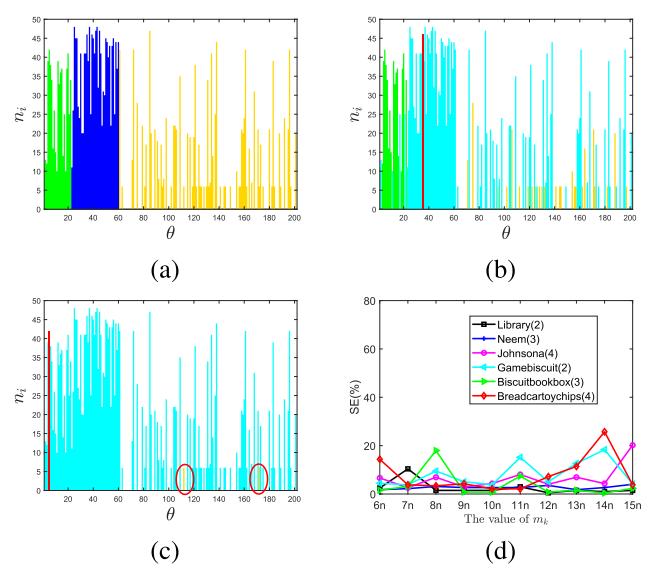


Fig. 6. The analysis of the parameter m_k . (a) The number of inliers of the generated model hypotheses. The bad model hypotheses are labeled in gold, and the others correspond to different model instances are labeled in different colors. (b) and (c) The results by using mutual information for model selection after the first and second iterations, respectively. The removed model hypotheses are labeled in cyan, and the estimated model instances are labeled in red. We also label the remaining model hypotheses in ellipse in (c). (d) The segmentation errors obtained by the proposed LGF fitting scheme with different values of m_k on six representative image pairs.

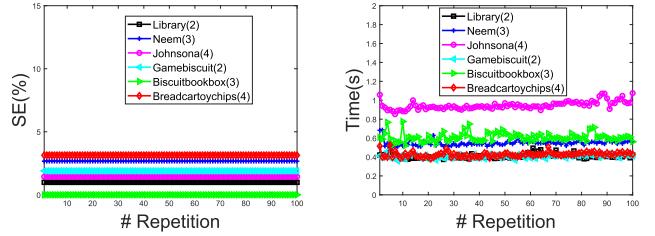


Fig. 7. Consistency analysis on six representative image pairs.

thus, the former requires more neighbors to support the corresponding preference sets. For the CPU time, it rises slowly with the increase of K value. Thus, we set the value of K to be larger than 12 for all the following experiments.

For the number of hash functions, LGF cannot obtain low segmentation errors if the value is too small; While it will require more CPU time if the value is too large; Therefore, we set b to $n/4$ where LGF obtains the lowest segmentation errors for most cases. For the value of ξ , we can see that LGF can obtain low segmentation errors when its value is larger than 0.7, since that it will filter out too many correspondences if ξ is too small, which leads to not enough significant sampling subsets for model fitting tasks. Thus, we set ξ to be larger than 0.7 for all the following experiments.

Next, we continue analyze the setting of the parameter m_k . Recalled that m_k has been used in Algorithms 1 and 2. Then, we first analyze m_k in Algorithm 2. m_k is used to judge if a model hypothesis is an effective model instance and the number of model instances are also derived from it. Thus, if we set m_k too large or too small, then we will wrongly estimate the number of model instances. In fact, it is not

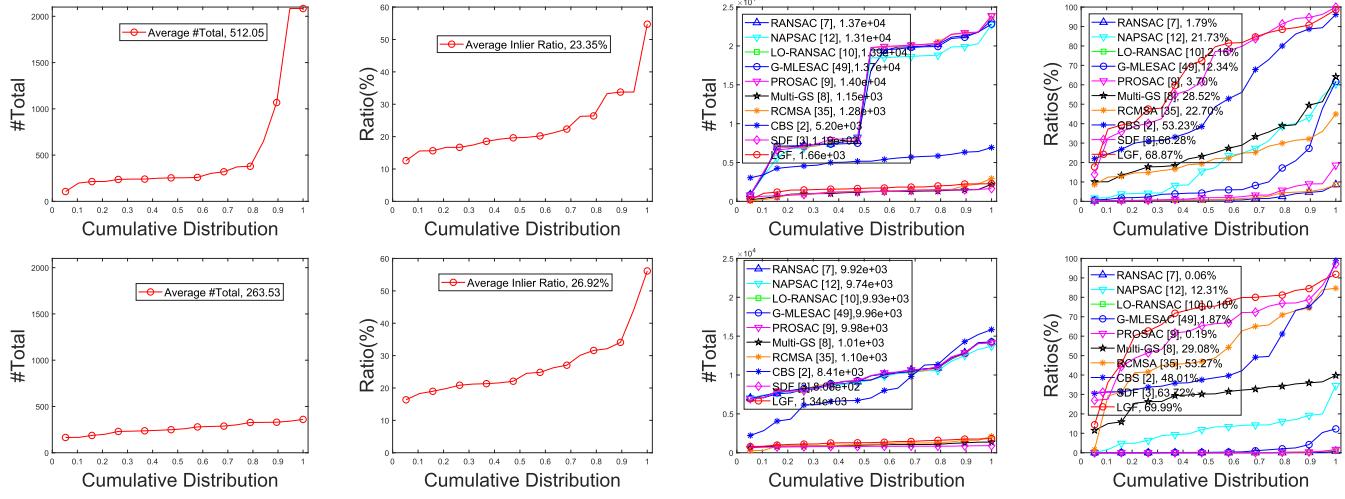


Fig. 8. Quantitative comparisons of night sampling methods on all image pairs from the AdelaideRMF dataset for homography-based segmentation (Top) and two-view-based motion segmentation (Bottom). (Left to Right) The total number of input correspondences, the average inlier ratios of input correspondences, the total number of sampled minimal subsets and the ratio of all-inlier minimal subsets on the sampled minimal subsets with respect to the cumulative distribution. A point on the curve with coordinate (x, y) denotes that there are $100 * x$ percent of image pairs which have values no more than y . The average value of each case is also shown in the legend.

hard to obtain a reasonable value of m_k . This is because a good model hypothesis generally has a large number of inliers. In contrast, for a bad model hypothesis, if it has a large number of inliers, then its mutual information is also large as well and it will be removed during the model selection process; Otherwise, it only has a small number of inliers.

We show an example in Fig. 6 for the “Library” image pair. From Figs. 6(b) and 6(c), we can see that, most of bad model hypotheses are removed during the model selection process, and the remaining model hypotheses only has a very small number of inliers. Thus, we have many options for m_k in Algorithm 2.

However, m_k will affect the quality of model hypotheses in Algorithm 1. We test different values of m_k for homography-based segmentation and two-view-based motion segmentation on six representative image pairs, and show the results in Fig. 6(d) (we do not show the CPU time since it has not obvious changes). We can see that, LGF obtains low segmentation errors for all six image pairs when $m_k = 9\% * n \sim 10\% * n$, where n is the number of input data points. Thus, we fix $m_k = 10\% * n$, which is also consistent with the setting in [2], [3], [24], for most image pairs.

2) *Consistency Analysis:* To test the consistency of the solutions provided by the proposed LGF, we repeat each experiment 100 times on the six representative image pairs for homography-based segmentation and two-view-based motion segmentation, and report the segmentation errors and the CPU time in Fig. 7.

We can see that, the segmentation errors obtained by LGF have not changes on all the six image pairs. For the CPU time, it varies only slightly for the repetitions. Therefore, LGF can provide consistent solutions for the two model fitting tasks.

3) *The Influence of the Components of LGF:* Recall that the proposed LGF is based on local-neighbor preservation and global-residual optimization. Thus, in this subsection, we analyze the influence of each component of LGF on the

TABLE I
PERFORMANCE OF THE THREE VERSIONS OF LGF FOR
HOMOGRAPHY-BASED SEGMENTATION (H) AND
TWO-VIEW-BASED MOTION SEGMENTATION (F)

	Method		LGF1	LGF2	LGF3
H	Library	SE(%)	1.98	1.48	0.99
		Time(s)	0.42	0.60	0.66
	Neem	SE(%)	2.66	2.66	2.66
		Time(s)	0.54	0.70	0.78
F	Johnsona	SE(%)	3.17	2.01	1.44
		Time(s)	0.91	1.03	1.12
	Gamebiscuit	SE(%)	14.55	1.89	1.89
		Time(s)	0.34	0.41	0.42
F	Biscuitbookbox	SE(%)	1.56	0.39	0.00
		Time(s)	0.52	0.61	0.63
	Breadcartoychips	SE(%)	13.45	20.62	3.13
		Time(s)	0.36	0.47	0.48

final fitting results. Note that, we adopt preference analysis and the min-hash technique to exploit local-neighbor preservation relationships, for sampling a small number of minimal subsets. Then we test a version (called as LGF1), which directly uses Euclidean distance to compute the initial elements of a minimal subset. That is, Eq. (4) is replaced by $\mathcal{M}\mathcal{S}_{x_i} = \{x_j | d_o(x_i, x_j) < \zeta, x_j \in X\}$, where $d_o(x_i, x_j)$ is the Euclidean distance between two correspondences. For the global-residual optimization technique, we define a novel weighting score, which considers the consensus of the elements in a minimal subset. Thus, we test a version (called as LGF2), which has not such consideration. After that, we also run the final version LGF3 as a baseline.

We test different versions of LGF for homography-based segmentation and two-view-based motion segmentation and we respectively test three image pairs from the AdelaideRMF dataset for each task. Then we report the segmentation errors and the CPU time in Table I.

We can see that, LGF3 obtains the minimal segmentation error among all three versions of LGF on all six image

TABLE II
PERFORMANCE OF THE TEN COMPETING SAMPLING METHODS FOR HOMOGRAPHY-BASED
SEGMENTATION (D1-D3) AND TWO-VIEW-BASED MOTION SEGMENTATION (D4-D6)

Method		RANSAC [8]	NAPSAC [13]	LO-RANSAC [11]	G-MLESAC [50]	PROSAC [10]	Multi-GS [9]	RCMSA [36]	CBS [2]	SDF [3]	LGF
D1	#Total	21333(0.46)	19872(3.88)	20861(0.46)	21137(3.69)	21380 (0.50)	1401(14.92)	1429(29.18)	5166(35.31)	1547(42.53)	2232(39.25)
	#S1	56(0.26)	587(2.95)	59(0.28)	563(2.66)	59(0.28)	163(11.63)	245(17.14)	1014 (19.63)	476(30.77)	564(25.27)
	#S2	43(0.20)	185(0.93)	37(0.18)	218(1.03)	48(0.22)	46(3.28)	172(12.04)	810 (15.68)	182(11.76)	312(13.98)
D2	#Total	20022(0.67)	18361(17.85)	20179 (0.68)	19312(5.80)	20069(0.84)	1316(30.55)	1309(15.51)	6330(38.50)	1463(55.02)	2250(59.56)
	#S1	92(0.46)	2165 (11.79)	96(0.48)	147(0.76)	108(0.54)	213(16.19)	67(5.12)	1623(25.64)	623(42.58)	670(29.78)
	#S2	14(0.07)	697(3.80)	14(0.07)	972 (5.03)	25(0.12)	56(4.26)	15(1.15)	19(0.30)	77(5.26)	360(16.00)
	#S3	28(0.14)	415(2.26)	28(0.14)	2(0.01)	36(0.18)	133(10.11)	121(9.24)	795 (12.56)	105(7.18)	310(13.78)
D3	#Total	7383(0.72)	6829(27.30)	7608 (0.99)	7335(1.06)	7427(1.23)	974(27.31)	879(29.92)	5784(48.67)	492(77.44)	2268(81.48)
	#S1	14(0.19)	486(7.12)	16(0.21)	29(0.40)	16(0.22)	58(5.95)	72(8.19)	1262 (21.82)	72(14.63)	497(21.91)
	#S2	30(0.41)	592(8.67)	49(0.64)	34(0.46)	61(0.82)	139(14.27)	48(5.46)	363(6.28)	183(37.20)	595(26.23)
	#S3	6(0.08)	246(3.60)	6(0.08)	3(0.04)	8(0.11)	21(2.16)	142 (16.15)	512 (8.85)	21(4.27)	301(13.27)
	#S4	3(0.04)	540(7.91)	4(0.05)	12(0.16)	6(0.08)	48(4.93)	1(0.11)	678 (11.72)	105(21.34)	455(20.06)
D4	#Total	7058(0.01)	7065 (8.83)	6939(0.00)	6957(4.20)	7084(0.37)	867(25.49)	889(73.12)	6140(30.52)	696(75.43)	1074(75.42)
	#S1	0(0.00)	47(0.67)	0(0.00)	292(4.20)	25(0.35)	104(12.00)	164(18.45)	799 (13.01)	216(31.03)	390(36.31)
	#S2	1(0.01)	577 (8.17)	0(0.00)	0(0.00)	1(0.01)	117(13.49)	486(54.67)	1075 (17.51)	309(44.40)	420(39.11)
D5	#Total	9891(0.00)	9342(14.28)	10018(0.00)	10016(0.44)	9924(0.02)	1005(34.23)	1104(48.82)	11280 (34.25)	800(66.88)	1280(91.88)
	#S1	0(0.00)	1016(10.88)	0(0.00)	0(0.00)	0(0.00)	174(17.31)	187(16.94)	1517 (13.45)	470(58.75)	672(52.50)
	#S2	0(0.00)	77(0.82)	0(0.00)	0(0.00)	0(0.00)	57(5.67)	196(17.75)	1001 (8.87)	15(1.88)	176(13.75)
	#S3	0(0.00)	241(2.58)	0(0.00)	44(0.44)	2(0.02)	113(11.24)	156(14.13)	1345 (11.92)	50(6.25)	328(25.62)
D6	#Total	10766(0.00)	10284(4.80)	10574(0.01)	10790(0.02)	10724(0.00)	1062(32.77)	1244(44.94)	14290 (31.35)	900(48.67)	1296(65.97)
	#S1	0(0.00)	39(0.38)	0(0.00)	0(0.00)	0(0.00)	66(6.21)	146(11.74)	992 (6.94)	18(2.00)	108(8.33)
	#S2	0(0.00)	3(0.03)	0(0.00)	0(0.00)	0(0.00)	17(1.60)	82(6.59)	885 (6.19)	0(0.00)	18(1.39)
	#S3	0(0.00)	178(1.73)	0(0.00)	2(0.02)	0(0.00)	98(9.23)	10(8.00)	1147 (8.03)	198(22.00)	396(30.56)
	#S4	0(0.00)	274(2.66)	1(0.01)	0(0.00)	0(0.00)	167(15.73)	321(25.80)	1456 (10.19)	222(24.67)	333(25.69)

Legend: #Total reports the total number of minimal subsets sampled within 5 seconds (and the ratio of all-inlier minimal subset and 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-Library; D2-Neem; D3-Johnsona; D4-Gamebiscuit; D5-Biscuitbookbox; D6-Breadcartoychips.)

pairs for the two model fitting tasks. Note that, two-view-based motion segmentation is a more challenging task than homography-based segmentation, since it is harder to sample an all-inlier minimal subset. Thus, LGF1 and LGF2 achieve similar segmentation errors as LGF3 for homography-based segmentation, while LGF1 fails to fit model instances on “Gamebiscuit” and “Breadcartoychips” image pairs and LGF2 fails on the “Breadcartoychips” image pair for two-view-based motion segmentation. For the CPU time, although LGF1 is slightly faster than LGF2 and LGF3 on all six image pairs, they have no significant differences. Therefore, the proposed local-neighbor preservation and global-residual optimization techniques can improve the performance of model fitting on the fitting accuracy without increasing significant time cost.

B. Results on Real Data

In this subsection, we evaluate the performance of the proposed LGF fitting scheme on the minimal subset sampling task for two popular fitting tasks, i.e., homography-based segmentation and two-view-based motion segmentation, and we also provide a quantitative comparison on all image pairs from the AdelaideRMF dataset with nine state-of-the-art sampling methods, including RANSAC, NAPSAC [13], LO-RANSAC [11], G-MLESAC [50], PROSAC [10], Multi-GS [9], RCMSA [36], CBS [2] and SDF [3]. After that, we also evaluate the performance of LGF on the final fitting task for the two popular fitting tasks, and provide a quantitative comparison with seven state-of-the-art fitting methods, including RansaCov [31], T-linkage [29], RPA [28], RCMSA [36], Prog-X [39], MSHF [35] and SDF [3].

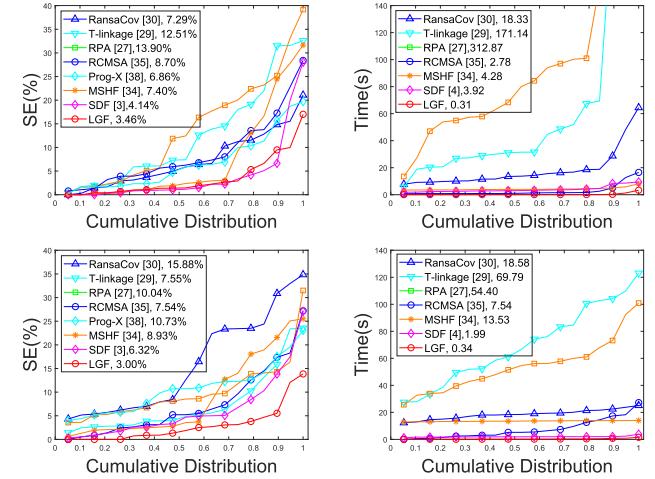


Fig. 9. Quantitative comparisons of seven fitting methods on all image pairs from the AdelaideRMF dataset for homography-based segmentation (Top) and two-view-based motion segmentation (Bottom). (Left to Right) The segmentation errors and the CPU time with respect to the cumulative distribution.

For the sampling minimal subsets, we report the details of the performance of the ten competing sampling methods on six image pairs for the two popular fitting tasks in Table II. We also report the quantitative comparisons of all competing sampling methods on all image pairs from AdelaideRMF dataset in Fig. 8. Here, it is wroth pointing out that, we can only compare the results obtained by all sampling methods within “reasonable” time. But, it is not an exact definition of “reasonable” time since it may change with problem complexity and processor. Then, we choose 5 seconds as an indicative reasonable time, for the two popular fitting tasks.

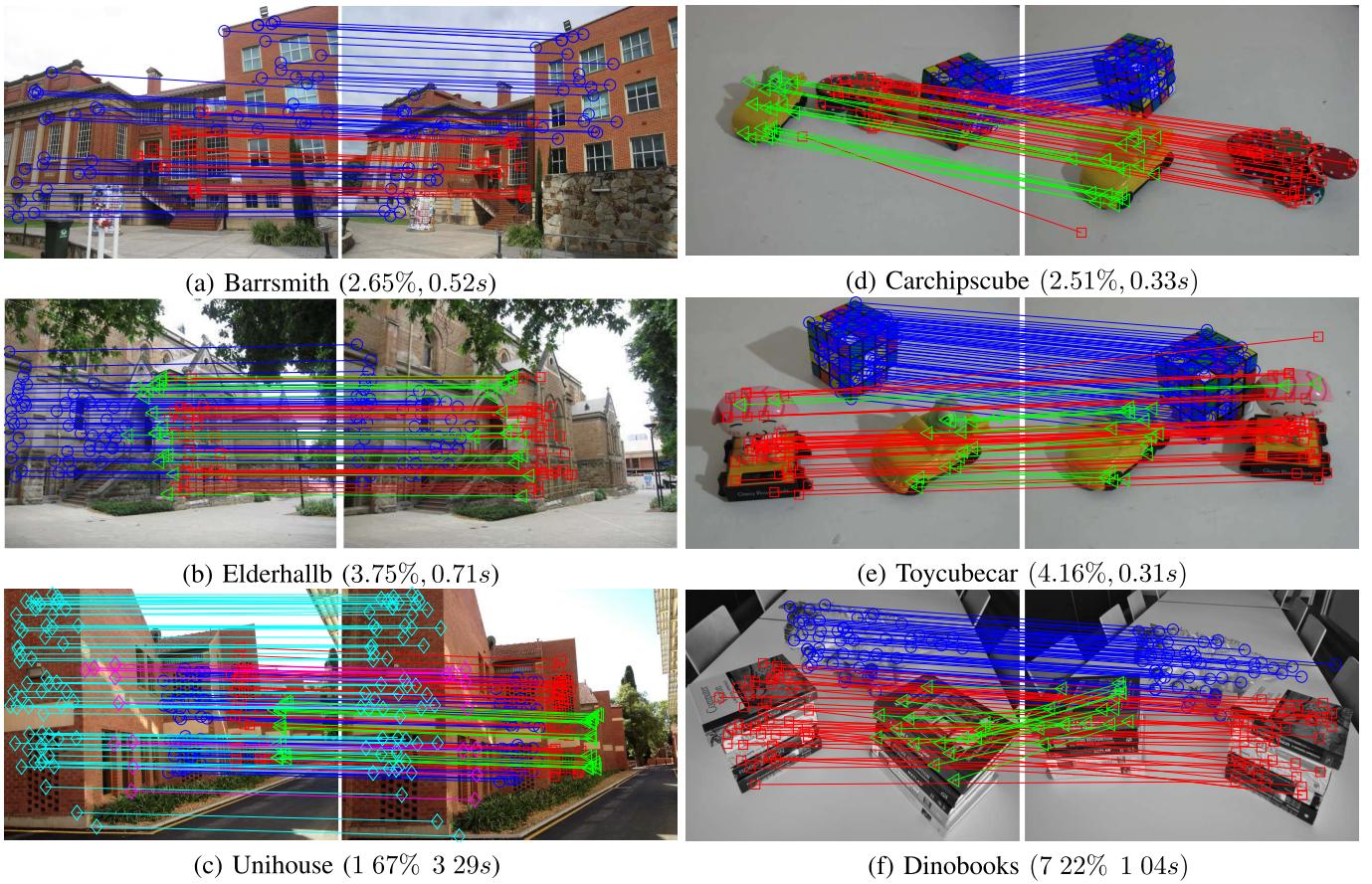


Fig. 10. Some challenging cases that the proposed LGF fitting method successfully estimates model instances ((a)-(c) for homography-based segmentation and (d)-(f) for two-view-based motion segmentation). For visibility, in the image pairs, at most 200 randomly selected inliers are presented. We also show the segmentation errors and the CPU time obtained by the proposed LGF fitting scheme.

For the final fitting results, we report the quantitative comparisons (the segmentation errors and the CPU time) of the eight competing fitting methods on all image pairs from AdelaideRMF dataset for the two popular fitting tasks in Fig. 9.

1) Homography-Based Segmentation: From Fig. 8, we can see that RANSAC, NAPSAC, LO-RANSAC, G-MLESAC and PROSAC sample an order of magnitude more minimal subsets than Multi-GS, RCMSA, CBS, SDF and LGF within 5 seconds for homography-based segmentation. This is because Multi-GS, RCMSA, CBS, SDF and LGF introduce more constraints, which require more time to deal with. However, only NAPSAC, Multi-GS, RCMSA, CBS, SDF and LGF achieve over 20% for the average ratio of all-inlier minimal subsets and all sampled minimal subsets, and RANSAC, LO-RANSAC and PROSAC cannot effectively sample high-quality minimal subsets. Note that, SDF and LGF have more larger value of the average ratio than the other seven competing sampling methods (LGF is the best one), and LGF also samples more number of minimal subsets than SDF within 5 seconds. From Table II, we can see that, LGF not only achieves the largest ratio of all-inlier minimal subset and the total number of subsets for two out of the three image pairs and the second largest ratio for another case, but also covers all model instances in data on a balanced manner for the three image pairs. That is, LGF provides high-quality sampled minimal

subsets, which will help improve the final fitting performance. It is worth pointing out that, a good sampling result consists of as more all-inlier minimal subsets as possible and as less impure minimal subsets as possible, since the impure minimal subsets will affect the final fitting performance. Thus, the ratio of all-inlier minimal subset and the total number of subsets can measure the sampling result more effectively than the total number.

From Fig. 9, we can see that LGF achieves the lowest average segmentation error among all eight fitting methods on all image pairs from AdelaideRMF dataset for the final fitting results. From the cumulative distribution, we also can see that LGF achieves low segmentation errors for most cases. More importantly, LGF has remarkably improved the computational efficiency, that is, LGF reduces the average CPU time from 2.78s, obtained by RCMSA (the fastest fitting method among the other seven competing fitting methods), to 0.31s. The reason behind this is that, LGF only deals with a small number of high-quality minimal subsets, which covers all model instances in data, and it also benefits from the effectiveness of the proposed model selection framework. It is worth pointing out that, only T-linkage, RCMSA, Prog-X, MSHF and LGF are able to automatically estimate the number of model instances in data, while the other three fitting methods require to provide the number of model instances in advance.

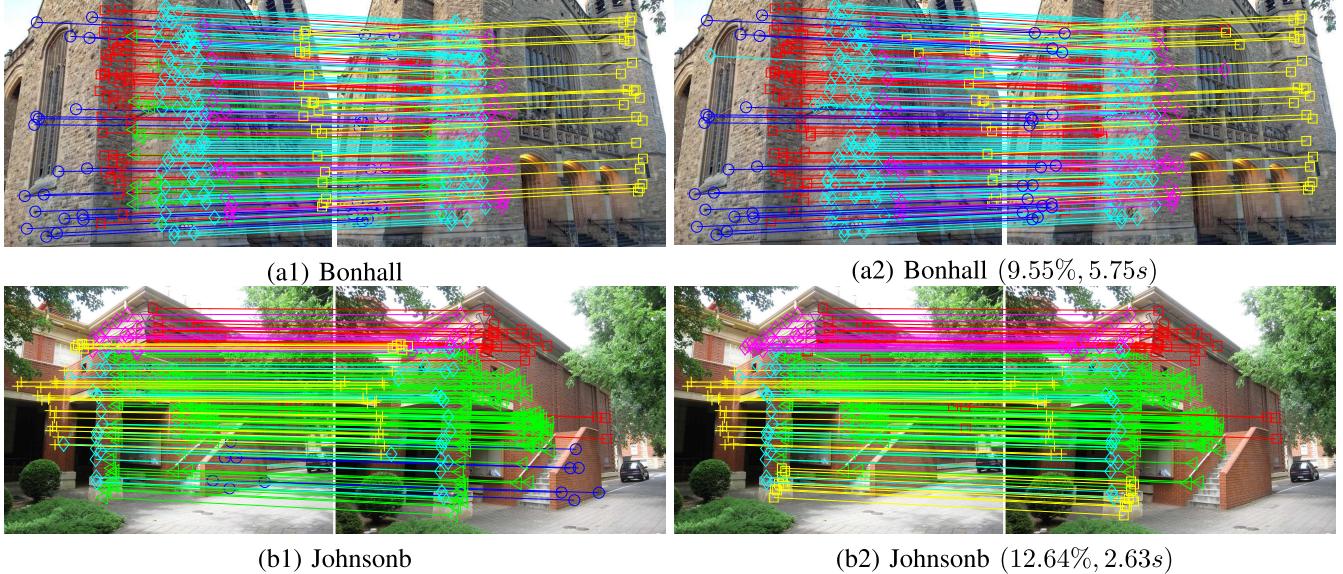


Fig. 11. Some challenging cases that the proposed LGF fitting scheme fails to estimate model instances for homography-based segmentation. (a1) and (b1) The ground truth results. (b1) and (b2) The fitting results obtained by LGF (we also show its segmentation errors and the CPU time).

2) *Two-View-Based Motion Segmentation*: From Fig. 8, we can see that RANSAC, NAPSAC, LO-RANSAC, G-MLESAC and PROSAC also sample more minimal subsets than other five sampling methods within 5 seconds for two-view-based motion segmentation, but only RCMSA, SDF and LGF achieve over 50% for the average ratio of all-inlier minimal subsets and all sampled minimal subsets. From the three cases in Table II, we also can see that LGF outperforms the other competing sampling methods for the performance on the ratio of all-inlier minimal subsets and all sampled minimal subsets, and it can cover all model instances in data even for the challenging case, i.e., the “Cubebreadtoychips” image pair.

As shown in Fig. 9, for the final fitting results, LGF is able to achieve low segmentation errors for most cases on all 19 image pairs within very few seconds. Compared with other seven fitting methods, LGF not only achieves the lowest average segmentation errors but also is the fastest fitting method. Thus, LGF is an effective and efficient fitting method.

VI. DISCUSSIONS

In this section, we further discuss the performance of the proposed LGF scheme. We show some challenging cases that LGF successfully estimates model instances in Fig. 10.

For the “Barrsmith” image pair, a model instance is split into two, which may cause high segmentation errors. However, LGF can avoid this situation due to the effectiveness of its global-residual optimization strategy. For the “Elderhallb” image pair, the three model instances have similar background, which may make it difficult to distinguish the input correspondences. While LGF is able to achieve low segmentation errors since LGF can preserve local-neighbor relationships of input correspondences. For the “Unihouse” image pair, the model instance (labeled carmine color in Fig. 10) only contains 4.17% inliers, while the other four model instances contain 23.99%, 23.80%, 23.99% and 7.49% inliers, respectively. As we know, most of state-of-the-art model fitting

methods cannot effectively deal with the “Unihouse” image pair. However, LGF not only correctly estimates the number of model instances but also achieves a low segmentation error within a few seconds, due to its high-quality sampled minimal subsets and the effective model selection framework.

For the two-view-based motion segmentation, the “Carchipscube” and “Toycubecar” image pairs also contain unbalanced number of inliers and LGF can obtain good performance on the fitting accuracy and computational efficiency. The model instances in the “Dinobooks” image pair have similar background and the case also contains a large proportion of outliers (i.e., 44.54%). Thus, it is hard to achieve a low segmentation error for a model fitting method. However, LGF is able to achieve the lowest segmentation errors among several state-of-the-art model fitting methods.

We also show some challenging cases that LGF fails to estimate model instances in Fig. 11. We can see that the two image pairs include extremely unbalanced number of inliers, i.e., the minimal and maximum inlier ratios are 6.20% and 31.41% for the “Bonhall” image pair, 2.43% and 46.84% for the “Johnsonb” image pair, respectively. Moreover, the correspondences in the “Bonhall” image pair are far part, which is a very challenging situation for both the minimal subset sampling and model selection. Note that, this weakness also affects most of state-of-the-art fitting methods.

VII. CONCLUSION

In this research, we propose a two-view approximately deterministic model fitting scheme (called LGF), which consists of an effective and efficient two-view approximately deterministic sampling algorithm and a simple yet effective model selection framework. For the proposed sampling algorithm, we exploit local-neighbor preservation relationships of input correspondences to roughly sample minimum subsets and then refine these subsets by a global-residual optimization strategy. The proposed sampling algorithm is able to sample a small number of high-quality minimal subsets, which include

a high ratio of all-inlier minimal subsets. For the proposed model selection framework, based on the effectiveness of our proposed weighting measure for each model hypothesis, we introduce the inlier noise scale constraint and the mutual information theory, to simultaneously estimate the number and the parameters of model instances within very few seconds. Compared with several state-of-the-art model fitting methods, extensive experiments on real images demonstrate the superior performance of LGF on both fitting accuracy and computational efficiency.

REFERENCES

- [1] L. Magri and A. Fusiello, "Fitting multiple heterogeneous models by multi-class cascaded T-linkage," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7460–7468.
- [2] R. Tennakoon, A. Sadri, R. Hoseinnezhad, and A. Bab-Hadiashar, "Effective sampling: Fast segmentation using robust geometric model fitting," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4182–4194, Sep. 2018.
- [3] G. Xiao, H. Wang, Y. Yan, and D. Suter, "Superpixel-guided two-view deterministic geometric model fitting," *Int. J. Comput. Vis.*, vol. 127, no. 4, pp. 323–339, Apr. 2019.
- [4] Z. L. Xun Xu and L.-F. Cheong, "3D rigid motion segmentation with mixed and unknown number of models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, no. 1, pp. 1–14, Jul. 2019.
- [5] J. Zhang *et al.*, "Learning two-view correspondences and geometry using order-aware network," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2019, pp. 5845–5854.
- [6] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Comput. Vis.*, to be published, doi: [10.1007/s11263-020-01359-2](https://doi.org/10.1007/s11263-020-01359-2).
- [7] C. Raposo, M. Antunes, and J. P. Barreto, "Piecewise-planar stereoscan: Sequential structure and motion using plane primitives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, no. 99, pp. 1–8, Aug. 2017.
- [8] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [9] T.-J. Chin, J. Yu, and D. Suter, "Accelerated hypothesis generation for multistructure data via preference analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 625–638, Apr. 2012.
- [10] O. Chum and J. Matas, "Matching with prosac—Progressive sample consensus," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 220–226.
- [11] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," in *Proc. Pattern Recognit.*, 2003, pp. 236–243.
- [12] Y. Kanazawa and H. Kawakami, "Detection of planar regions with uncalibrated stereo using distributions of feature points," in *Proc. Brit. Mach. Vis. Conf.*, 2004, pp. 2701–2710.
- [13] D. R. Myatt, P. H. S. Torr, S. J. Nasuto, J. M. Bishop, and R. Craddock, "NAPSAC: High noise, high dimensional robust estimation—it's in the bag," in *Proc. Brit. Mach. Vis. Conf.*, 2002, pp. 458–467.
- [14] G. Xiao, H. Wang, Y. Yan, and D. Suter, "Superpixel-based two-view deterministic fitting for multiple-structure data," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 517–533.
- [15] H. Li, "Consensus set maximization with guaranteed global optimality for robust geometry estimation," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 1074–1080.
- [16] K. H. Lee and S. W. Lee, "Deterministic fitting of multiple structures using iterative MaxFS with inlier scale estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 41–48.
- [17] R. Litman, S. Korman, A. Bronstein, and S. Avidan, "Inverting RANSAC: Global model detection via inlier rate estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5243–5251.
- [18] J. Fredriksson, V. Larsson, and C. Olsson, "Practical robust two-view translation estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2684–2690.
- [19] T.-J. Chin, P. Purkait, A. Eriksson, and D. Suter, "Efficient globally optimal consensus maximisation with tree search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 758–772, Apr. 2017.
- [20] H. Le, T.-J. Chin, A. Eriksson, T.-T. Do, and D. Suter, "Deterministic approximate methods for maximum consensus robust fitting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, no. 1, pp. 1–14, 2019.
- [21] A. Z. Broder, M. Charikar, A. M. Frieze, and M. Mitzenmacher, "Min-wise independent permutations," *J. Comput. Syst. Sci.*, vol. 60, no. 3, pp. 630–659, Jun. 2000.
- [22] P. Indyk and R. Motwani, "Approximate nearest neighbors: Towards removing the curse of dimensionality," in *Proc. 13th Annu. ACM Symp. Theory Comput.*, 1998, pp. 604–613.
- [23] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul./Oct. 1948.
- [24] R. B. Tennakoon, A. Bab-Hadiashar, Z. Cao, R. Hoseinnezhad, and D. Suter, "Robust model fitting using higher than minimal subset sampling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 350–362, Feb. 2016.
- [25] F. Wen, R. Ying, Z. Gong, and P. Liu, "Efficient algorithms for maximum consensus robust fitting," *IEEE Trans. Robot.*, vol. 36, no. 1, pp. 92–106, Feb. 2020.
- [26] P. Purkait, T.-J. Chin, A. Sadri, and D. Suter, "Clustering with hypergraphs: The case for large hyperedges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 9, pp. 1697–1711, Sep. 2017.
- [27] G. Xiao, H. Wang, T. Lai, and D. Suter, "Hypergraph modelling for geometric model fitting," *Pattern Recognit.*, vol. 60, no. 1, pp. 748–760, Dec. 2016.
- [28] L. Magri and A. Fusiello, "Multiple structure recovery via robust preference analysis," *Image Vis. Comput.*, vol. 67, pp. 1–15, Nov. 2017.
- [29] R. Toldo and A. Fusiello, "Robust multiple structures estimation with J-linkage," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 537–547.
- [30] L. Magri and A. Fusiello, "T-linkage: A continuous relaxation of J-linkage for multi-model fitting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3954–3961.
- [31] L. Magri and A. Fusiello, "Multiple models fitting as a set coverage problem," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 3318–3326.
- [32] L. Xu, E. Oja, and P. Kultanen, "A new curve detection method: Randomized Hough transform (RHT)," *Pattern Recognit. Lett.*, vol. 11, no. 5, pp. 331–338, May 1990.
- [33] S. Mittal, S. Anand, and P. Meer, "Generalized projection-based M-estimator," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 12, pp. 2351–2364, Dec. 2012.
- [34] H. Wang, T.-J. Chin, and D. Suter, "Simultaneously fitting and segmenting multiple-structure data with outliers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1177–1192, Jun. 2012.
- [35] H. Wang, G. Xiao, Y. Yan, and D. Suter, "Searching for representative modes on hypergraphs for robust geometric model fitting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 697–711, Mar. 2019.
- [36] T. T. Pham, T.-J. Chin, J. Yu, and D. Suter, "The random cluster model for robust geometric fitting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1658–1671, Aug. 2014.
- [37] X. Zhao, Y. Zhang, and B. Luo, "Energy minimization based alternate sampling and clustering for geometric model fitting," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2019, pp. 1570–1574.
- [38] P. Amayo, P. Pinies, L. M. Paz, and P. Newman, "Geometric multi-model fitting with a convex relaxation algorithm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8138–8146.
- [39] D. Barath and J. Matas, "Progressive-X: Efficient, anytime, multi-model fitting algorithm," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2019, pp. 3779–3787.
- [40] Y. Zheng and D. Doermann, "Robust point matching for nonrigid shapes by preserving local neighborhood structures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 643–649, Apr. 2006.
- [41] M. S. Charikar, "Similarity estimation techniques from rounding algorithms," in *Proc. 13th Annu. ACM Symp. Theory Comput.*, 2002, pp. 380–388.
- [42] D. Gorisse, M. Cord, and F. Precioso, "Locality-sensitive hashing for Chi2 distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 2, pp. 402–409, Feb. 2012.
- [43] L. Magri and A. Fusiello, "Reconstruction of interior walls from point cloud data with min-hashed J-Linkage," in *Proc. Int. Conf. 3D Vis. (3DV)*, Sep. 2018, pp. 131–139.
- [44] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, May 2019.
- [45] K.-M. Lee, P. Meer, and R.-H. Park, "Robust adaptive segmentation of range images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 2, pp. 200–205, 1998.
- [46] A. Bab-Hadiashar and D. Suter, "Robust segmentation of visual data using ranked unbiased scale estimate," *Robotica*, vol. 17, no. 6, pp. 649–660, Nov. 1999.

- [47] H. Wang and D. Suter, "Robust fitting by adaptive-scale residual consensus," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 107–118.
- [48] H. S. Wong, T.-J. Chin, J. Yu, and D. Suter, "Dynamic and hierarchical multi-structure geometric model fitting," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1044–1051.
- [49] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [50] B. J. Tordoff and D. W. Murray, "Guided-MLESAC: Faster image transform estimation by using matching priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1523–1535, Oct. 2005.



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 to 2018, he was a Postdoctoral Fellow with the School of Aerospace Engineering, Xiamen University, China. He is currently a Full Professor with Minjiang University, China. He has published over 30 papers in the international journals and conferences, including *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, *International Journal of Computer Vision*, *Pattern Recognition*, *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, *ICCV*, *ECCV*, *ACCV*, *AAAI*, *ICIP*, *ICARCV*, and so on. His research interests include machine learning, computer vision, pattern recognition, and bioinformatics. He has been awarded the Best PhD Thesis in Fujian Province and the Best PhD Thesis Award in the China Society of Image and Graphics (a total of ten winners in China). He has also served on the program committee (PC) of *CVPR*, *ICCV*, *ECCV*, *AAAI*, and so on. He was the General Chair of *IEEE BDCloud 2019*.

PATTERN ANALYSIS AND MACHINE INTELLIGENCE, *International Journal of Computer Vision*, *Pattern Recognition*, *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, *ICCV*, *ECCV*, *ACCV*, *AAAI*, *ICIP*, *ICARCV*, and so on. His research interests include machine learning, computer vision, pattern recognition, and bioinformatics. He has been awarded the Best PhD Thesis in Fujian Province and the Best PhD Thesis Award in the China Society of Image and Graphics (a total of ten winners in China). He has also served on the program committee (PC) of *CVPR*, *ICCV*, *ECCV*, *AAAI*, and so on. He was the General Chair of *IEEE BDCloud 2019*.



Jiayi Ma (Member, IEEE) 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, Los Angeles, CA, USA. He is currently a Professor with the Electronic Information School, Wuhan University. He has authored or coauthored more than 120 refereed journal and conference papers, including *IEEE TPAMI/TIP*, *IJCIV*, *CVPR*, *ICCV*, and so on. His research interests include computer vision, machine learning, and pattern recognition.

Dr. Ma has been identified in the 2019 Highly Cited Researchers list from the Web of Science Group. He was a recipient of the Natural Science Award of Hubei Province (first class), the CAAI (Chinese Association for Artificial Intelligence) Excellent Doctoral Dissertation Award (a total of eight winners in China), and the Chinese Association of Automation (CAA) Excellent Doctoral Dissertation Award (a total of ten winners in China). He is an Editorial Board Member of *Information Fusion* and *Neurocomputing*.



Shiping Wang (Member, IEEE) 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.



Changwen Chen (Fellow, IEEE) received the B.S. degree from the University of Science and Technology of China in 1983, the M.S.E.E. degree from the University of Southern California in 1986, and the Ph.D. degree from the University of Illinois at Urbana-Champaign in 1992.

He was with the Faculty of Electrical and Computer Engineering, University of Rochester, from 1992 to 1996, and also with the Faculty of Electrical and Computer Engineering, University of Missouri, Columbia, MO, USA, from 1996 to 2003. He was the Allen Henry Endow Chair Professor with the Florida Institute of Technology from 2003 to 2007. Since 2008, he has been an Empire Innovation Professor of Computer Science and Engineering with the University at Buffalo, The State University of New York. He is currently the Dean of the School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen.

Dr. Chen has been an SPIE Fellow since 2007. He and his students received nine best paper awards or best student paper awards over the past two decades. He also received several research and professional achievement awards, including the Sigma Xi Excellence in Graduate Research Mentoring Award in 2003, the Alexander von Humboldt Research Award in 2009, the University at Buffalo Exceptional Scholar Sustained Achievement Award in 2012, and the State University of New York System Chancellors Award for Excellence in Scholarship and Creative Activities in 2016. His research is supported by NSF, DARPA, Air Force, NASA, Whitaker Foundation, Microsoft, Intel, Kodak, Huawei, and Technicolor. He was the Editor-in-Chief of the *IEEE TRANSACTIONS ON MULTIMEDIA* from 2014 to 2016. He has served as the Editor-in-Chief for the *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY* from 2006 to 2009. He is an Editor of several other major IEEE Transactions and journals, including the *Proceedings of the IEEE*, the *IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS*, and the *IEEE JOURNAL ON EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS*. He has served as the Conference Chair for several major IEEE, ACM, and SPIE conferences related to multimedia video communications and signal processing.