Accelerated Guided Sampling for Multi-Structure Model Fitting

Taotao Lai, Hanzi Wang*, *Senior Member, IEEE,* Yan Yan, *Member, IEEE,* Tat-Jun Chin, *Member, IEEE,* Jin Zheng, *Member, IEEE,* and Bo Li

***Abstract*—The performance of many robust model fitting techniques is largely dependent on the quality of the generated hypotheses. In this paper, we propose a novel guided sampling method, called Accelerated Guided Sampling (AGS), to efficiently generate accurate hypotheses for multi-structure model fitting. Based on the observations that residual sorting can effectively reveal data relationship (i.e., determine whether two data points belong to the same structure), and keypoint matching scores can be used to distinguish inliers from gross outliers, AGS effectively combines the benefits of residual sorting and keypoint matching scores to efficiently generate accurate hypotheses via information theoretic principles. Moreover, we reduce the computational cost of residual sorting in AGS by designing a new residual sorting strategy, which only sorts the top ranked residuals of input data, rather than all input data. Experimental results demonstrate the effectiveness of the proposed method in computer vision tasks such as homography matrix and fundamental matrix estimation.**

***Index Terms*—Robust model fitting, hypothesis generation, residual sorting, keypoint matching scores, multiple structures.**

I. Introduction

ROBUST model fitting plays an important role in computer vision, and has been widely employed in 3-D data segmentation [1], object recognition [2], automated rebar detection [3], visual tracking [4], indoor navigation [5], [6], and so on. Generating a set of accurate hypotheses is critical to the success of most model fitting methods such as [1], [7]–[11]. Many recently proposed model fitting methods (e.g., [8], [11]) use random sampling to generate model hypotheses. However, the drawback of random sampling is that, the probability of sampling an all-inlier *minimal subset* exponentially reduces when the order of the geometric model increases or the inlier ratio of *input data* decreases [12]. In this paper, we focus on guided sampling for model fitting based on image pairs, and input data are candidate image correspondences. A minimal subset represents a subset of input data with size m, where m is the minimum number of image correspondences required to estimate the parameters of a geometric model. Therefore, the computational cost of random sampling to generate an accurate hypothesis close to the true model is expensive.

To improve the efficiency of hypothesis generation, some sampling methods (e.g., [13]–[15]) have been proposed to guide the sampling process by using prior probabilities derived from (keypoint) matching scores. These methods can rapidly generate accurate hypotheses for single-structure data (even with more than 90% outliers). However, these methods may not obtain *one clean solution* for multi-structure data (i.e., at least one all-inlier or clean minimal subset is sampled for each model instance) within *a reasonable time* due to the fact that these methods usually sample cross-structure minimal subsets, especially for high-order geometric models [16]. In such a case, data in cross-structure minimal subsets may have high matching scores, but they are from different model instances. Here, we respectively choose 5 seconds and 10 seconds as indicative "*reasonable time*", for homography matrix and fundamental matrix estimation as done in [17].

Several guided sampling methods (e.g., [16], [18]) have been proposed to accelerate accurate hypothesis generation for two-stage fitting methods (e.g., [11], [19]–[21]). These two-stage model fitting methods first generate a number of hypotheses, and then perform model selection. The two steps are performed disjointedly. These fitting methods will break down if there is no single clean solution obtained in the hypothesis generation process. Moreover, it is time-consuming for these guided sampling methods to achieve a clean solution for multi-structure data with high outlier rates, because the sampling weights are computationally expensive. Note that, in multi-structure data, the outliers of one structure may contain both gross outliers and pseudo-outliers (i.e., the inliers belonging to one structure are the outliers to the other structures).

In this paper, we propose a new hypothesis generation method, named Accelerated Guided Sampling (AGS), which aims to efficiently generate accurate hypotheses. Different from model fitting methods such as USAC [9], which includes all important steps of fitting process, such as hypothesis generation and degeneracy detection, whereas AGS mainly focuses on improving the efficiency of hypothesis generation and it can be used in other fitting methods such as USAC. In more detail, the basic steps of AGS are given as follows.

During the hypothesis generation process, AGS first computes data preference correlations (i.e., conditional probabilities) by residual sorting as done in [16], [18]. Based on the observation that correlations between the inliers from the same structure have higher values [16], we propose to use

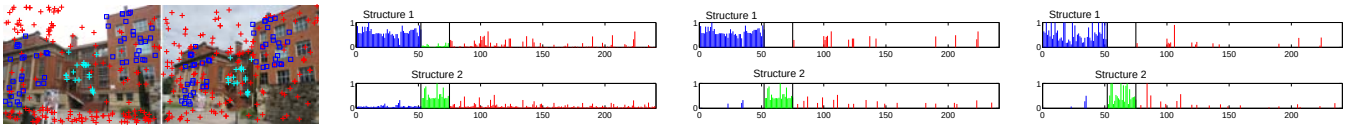
Taotao Lai, Hanzi Wang and Yan Yan are with the Fujian Key Laboratory of Sensing and Computing for Smart City, School of Information Science and Engineering, Xiamen University, Xiamen 361005, China (e-mail: laitaotao@gmail.com, hanzi.wang@xmu.edu.cn, yanyan@xmu.edu.cn).

Taotao Lai is also with the Digital Fujian Institute of Big Data for Agriculture and Forestry, College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China.

Tat-Jun Chin is with the ACVT and School of Computer Science, The University of Adelaide, Adelaide, SA 5005, Australia (e-mail: tat-jun.chin@adelaide.edu.au).

Jin Zheng and Bo Li are with Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China (e-mail: JinZheng@buaa.edu.cn, boli@buaa.edu.cn).

*Corresponding author.



(a) (b) (c) (d)

Fig. 1. Illustration of the proposed AGS for homography matrix estimation. (a) The *Barrsmith* image pair from the AdelaideRMF dataset with two model instances. (b) The correlation values of the input data with an inlier (the bold black vertical lines separating the structures and the outliers). (c) The correlations of the significant data selected by AGS based on information theoretic principles. The data with low correlation values are filtered out. (d) The sampling weights computed by using both the correlations of (c) and the correspondence matching scores. Compared with (c), the influence of gross outliers is reduced.

information theoretic principles to select significant data (that are more likely from the same model instance) with larger correlation values. On the other hand, the matching scores of inliers usually have higher values than those of outliers. To reduce the influence of gross outliers, AGS computes the sampling weights by using both the correlations and matching scores. Moreover, to increase the efficiency of residual sorting in AGS, we propose a new residual sorting strategy, which significantly reduces the computational cost compared with the strategy used in [16]. Finally, AGS samples data to generate a model hypothesis by using the sampling weights from the selected significant data.

Fig. 1 illustrates the main steps of AGS. In Fig. 1(a), the gross outliers are marked with red crosses, while the inliers of the two structures are marked with different colors, respectively. In Fig. 1(b) to Fig. 1(d), the x-axis means the rearranged data index grouped according to the structure membership, and the y-axis means the corresponding correlation values (Fig. 1(b) and (c)) or the sampling weights (Fig. 1(d)). First, the correlation values between an input datum $\boldsymbol{x}_i$ and input data are computed. The correlations between $\boldsymbol{x}_i$ and the data from the same structure that $\boldsymbol{x}_i$ belongs to, usually have large values (see Fig. 1(b), where $\boldsymbol{x}_i$ is an inlier belonging to the group drawn in the blue lines (the top row) or in the green lines (the bottom row)). Then, AGS uses information theoretic principles to filter out the data whose correlation values with regard to $\boldsymbol{x}_i$ are low (see Fig. 1(c)). As a result, most of the retained data are from the same structure. At last, AGS computes sampling weights (see Fig. 1(d)) by using both the correlations of the retained data in Fig. 1(c) and the corresponding matching scores, by which the influence of gross outliers is reduced. AGS samples minimal subsets from the retained data with the sampling weights, by which it increases the probability of sampling clean minimal subsets.

Compared with MultiGS [16] and ModeSamp [18], AGS has two main advantages: (1) AGS offers substantial speed-ups for hypothesis generation. To sample a minimal subset, MultiGS and ModeSamp compute correlations by sorting residuals of all input data $m-1$ times (e.g., $m=7$ or 8 for fundamental matrix estimation). However, AGS computes correlations only once. Thus, the computational cost of AGS to perform correlation computation is approximately equal to $1/(m-1)$ of that of MultiGS and ModeSamp for hypothesis generation. (2) Different from MultiGS and ModeSamp, which sample minimal subsets from all input data, AGS samples minimal subsets from the selected significant data. Thus, AGS significantly increases the probability of sampling all-inlier minimal subsets.

In summary, the main contribution of this work is that we develop an efficient guided sampling method for model fitting. More specifically, (1) we propose an effective hypothesis generation strategy, which uses information theoretic principles to select significant data from input data based on data correlations. Moreover, to reduce the influence of gross outliers, we compute the sampling weights of data by using both data correlations and matching scores. AGS samples a minimal subset from the selected significant data with the sampling weights, which leads to high computational efficiency (i.e., data correlations are computed only once for sampling a minimal subset) and high sampling effectiveness. (2) We propose an efficient residual sorting strategy, which effectively takes advantage of the information obtained by previous sorting results to reduce computational cost.

The rest of the paper is organized as follows: In Section II, we review related work. In Section III, we present the proposed AGS method. In Section IV, we show experimental results on both homography matrix and fundamental matrix estimation. In Section V, we draw the conclusions.

II. Related Work

A large number of sampling methods (e.g., [14]–[16], [22], [23]) have been proposed to improve the efficiency of hypothesis generation. Several methods assume that inliers from the same structure tend to be spatially close in the image domain. For example, NAPSAC [22] and Proximity [24] concentrate on sampling neighboring data. GroupSAC [25] first partitions input data into a number of groups through clustering or segmentation, and then guides the sampling process by using the information obtained from the groups. However, the assumption of these methods may be violated in some cases. For example, in Fig. 1(a), the inliers of the structure marked in blue are separated by the other structure marked in cyan, which causes that the inliers of the former structure are not spatially neighboring. Moreover, the solutions obtained by these methods may be inaccurate because the spans of the minimal subsets sampled by these methods are relatively small [26].

[23] embeds an inner RANSAC in the standard RANSAC to perform local exploration for accelerating accurate hypothesis generation. However, for data with a large number of inliers, the inner RANSAC may increase the computational time by an order of magnitude. To overcome this disadvantage, LO$^+$−RANSAC [27] proposes to limit the number of inliers, which are used to estimate model parameters in the inner

RANSAC. [28], [29] extends the idea of local exploration, and it iteratively performs local exploration until a local convergence result is obtained. However, for data with high outlier rates and high-order geometric models, these methods cannot sample accurate hypotheses efficiently.

In [13]–[15], hypothesis generation is guided by using information exploited from matching scores. Both PROSAC [13] and Guided-MLESAC [14] focus on sampling minimal subsets from better matched data. EVSAC [15] guides the sampling process by using conditional inlier probabilities derived from the closest-matching scores. These methods can efficiently tackle single-structure data. However, within a reasonable time, these methods usually cannot achieve one clean solution for multi-structure data, because the data (with high matching scores) in a minimal subset sampled by these methods are probably from different structures as pointed out in [16], [18].

MultiGS [16] and ModeSamp [18] are two recently proposed guided sampling methods, which use the information derived from residual sorting, and both can handle multi-structure data. ModeSamp is an improved version of MultiGS. ModeSamp improves the seed datum (i.e., the first sampled datum of a minimal subset) selection strategy of MultiGS: instead of randomly selecting the seed datum from all input data, ModeSamp selects the seed datum by performing mode seeking over permutations. However, the computation of sampling weights in both MultiGS and ModeSamp is time-consuming. Thus, these methods may be incapable of sampling accurate hypotheses for the structures with high outlier rates. IMaxFS-ISE [30] also takes advantage of residual information to guide sampling for multi-structure data by performing a sequential "fit-and-remove" procedure. Nevertheless, as noted by [11], [31], this method may suffer from the disadvantage of the sequential "fit-and-remove" procedure (i.e., inaccurately estimating the parameters of one structure may lead to incorrect estimation of the parameters of the remaining structures).

III. Accelerated Guided Sampling

In this section, we first introduce the notation used in this paper. Then we present an effective guided sampling strategy. To further improve the performance of AGS, we also propose a new residual sorting strategy. At last, we describe the complete AGS method.

A. Notation

Let $\boldsymbol{\mathcal{X}} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$ denote a set of N input data (i.e., the candidate image correspondences between an input image pair), and $\boldsymbol{\psi} = \{\psi_1, \psi_2, \ldots, \psi_N\}$ be the corresponding (keypoint) matching scores (assuming that a similarity metric is used to establish putative matches). A tentative hypothesis set $\boldsymbol{\Theta} = \{\theta_1, \theta_2, \ldots, \theta_c\}$ is estimated from a number of sampled minimal subsets. For the ith data $\boldsymbol{x}_i$, the residual vector $\mathbf{r}^i$ of $\boldsymbol{x}_i$ with regard to the generated c hypotheses is written as

$$\mathbf{r}^i = [r_1^i,\ r_2^i, \ldots, r_c^i]. \tag{1}$$

The permutation

$$\boldsymbol{\kappa}^i = [\kappa_1^i,\ \kappa_2^i, \ldots, \kappa_c^i] \tag{2}$$

of residual indices is obtained by sorting $\mathbf{r}^i$ in a non-descending order. $\boldsymbol{\kappa}^i$ is the preference of $\boldsymbol{x}_i$ to the c hypotheses, where $u, v \in \{1, 2, \ldots, c\}$, and $u < v \Longrightarrow r^i_{\kappa^i_u} \leq r^i_{\kappa^i_v}$.

$\boldsymbol{\kappa}^i_{1:w}$ is a vector including the first-w elements of $\boldsymbol{\kappa}^i$, where w is the window size. Here, $1 \leq w \leq c$ and w is set to $w = [\alpha * c]$, where $0 < \alpha < 1$ and the notation $[\cdot]$ means rounding the value. We compute the correlation $d_{i,j}$ of the ith data point $\boldsymbol{x}_i$ with regard to the jth data point $\boldsymbol{x}_j$ as [16]:

$$d_{i,j} = \frac{1}{w}\ |\boldsymbol{\kappa}^i_{1:w} \cap \boldsymbol{\kappa}^j_{1:w}|, \tag{3}$$

where $|\boldsymbol{\kappa}^i_{1:w} \cap \boldsymbol{\kappa}^j_{1:w}|$ represents the number of the same elements shared by $\boldsymbol{\kappa}^i_{1:w}$ and $\boldsymbol{\kappa}^j_{1:w}$. The correlation between any two inliers from the same structure usually have larger value than those from different structures, as shown in Fig. 1(b).

B. The Guided Sampling Strategy

To generate accurate hypotheses, AGS first computes the correlations (between input data and a randomly selected seed datum) by Eq. (3). Then, AGS selects significant data by using information theoretic principles based on the computed correlations, and computes the sampling weights of the selected significant data based on both the correlations and matching scores. At last, AGS samples a minimal subset by using the computed sampling weights from the selected significant data.

More specifically, if c hypotheses have been generated, to generate the $(c+1)$-th hypothesis by the proposed guided sampling strategy, a minimal subset is sampled as follows. A seed datum $\boldsymbol{x}_{s_1}$ is randomly chosen from the input data $\boldsymbol{\mathcal{X}}$, where s_1 is an data index and $s_1 \in \{1, 2, \ldots, N\}$. Here, N is the number of the input data $\boldsymbol{\mathcal{X}}$. Then, the correlation $d_{s_1,j}$ between $\boldsymbol{x}_{s_1}$ and the jth datum $\boldsymbol{x}_j$ of $\boldsymbol{\mathcal{X}}$ is computed by Eq. (3), where $j \in \{1, 2, \ldots, N\}$. Considering s_1 is fixed, we can set $D_j \equiv d_{s_1,j}$ for simplification, yielding a correlation vector

$$\mathbf{D} = [D_1,\ D_2, \ldots, D_N]. \tag{4}$$

As in [11], the gap ϕ_j between the maximum of $\mathbf{D}^\delta$ and the δ-th power of D_j is defined as

$$\phi_j = \max(\mathbf{D}^\delta) - D_j^\delta. \tag{5}$$

The gap ϕ_j computed by Eq. (5) is guaranteed to be no less than zero and is used as a complementary value, which is used to calculate an entropy threshold for effectively selecting significant data with larger correlation values. The lower the value of ϕ_j is, the more likely the jth data and $\boldsymbol{x}_{s_1}$ are from the same structure. This is because that, in Eq. (3), the correlation will have a high value when the jth data and $\boldsymbol{x}_{s_1}$ belong to the same structure. The influence of δ will be evaluated in Section IV-D.

The probability of ϕ_j is defined as

$$\eta(\phi_j) = \phi_j \Big/ \sum_{k=1}^{N} \phi_k. \tag{6}$$

According to [32], the quantity of information provided by the jth correlation is measured as

$$e_j = -\log(\eta(\phi_j) + \epsilon), \tag{7}$$

where ϵ is an infinitesimal positive number. The entropy of $\mathbf{D}$ is computed as

$$E = \sum_{j=1}^{N} \eta(\phi_j) e_j. \tag{8}$$

AGS selects significant data as follow:

$$\boldsymbol{\mathcal{X}}^* = \{\boldsymbol{x}_j | E < e_j\}. \tag{9}$$

According to Eqs. (5)–(9), the data with a lower quantity of information than E are rejected. The selected significant data have a larger quantity of information, and are more likely from the same structure as $\boldsymbol{x}_{s_1}$ (as shown in Fig. 1(c)). Sampling data subsets from the selected significant data can increase the probability of sampling clean data subsets. A set of indices

$$\{n_1, n_2, \ldots, n_L\} \subset \{1, 2, \ldots, N\}, \tag{10}$$

is found to identify the elements of the input data $\boldsymbol{\mathcal{X}}$, which are present in $\boldsymbol{\mathcal{X}}^*$, where L is the number of $\boldsymbol{\mathcal{X}}^*$.

We note that some model fitting methods [11], [33] also utilize information theoretic principles. However, the use of information theoretic principles in the proposed method is significantly different: in each *hypothesis generation process*, based on the computed preference correlations and information theoretic principles, the proposed method tries to choose the significant data, arising from the same structure that the chosen seed datum belongs to, for guided sampling. In other words, we use information theoretic principles to *guide the sampling process*. In contrast, [33] employs information theoretic principles to filter out gross outliers in the *data pre-processing step*, while [11] uses information theoretic principles to select significant hypotheses in the *model selection step*.

After selecting significant data, AGS chooses the correlation values $\mathbf{D}^*$ between the selected significant data $\boldsymbol{\mathcal{X}}^*$ and $\boldsymbol{x}_{s_1}$ from $\mathbf{D}$, and the correspondence matching scores $\boldsymbol{\psi}^*$ of $\boldsymbol{\mathcal{X}}^*$ from $\boldsymbol{\psi}$, where $\mathbf{D}^* = [D_{n_1}, D_{n_2}, ..., D_{n_L}]$ and $\boldsymbol{\psi}^* = [\psi_{n_1}, \psi_{n_2}, ..., \psi_{n_L}]$. The selected correlations and matching scores are respectively normalized as $\mu_l = D_{n_l} \Big/ \sum_{k=1}^{L} D_{n_k}$ and $\nu_l = \psi_{n_l} \Big/ \sum_{k=1}^{L} \psi_{n_k}$, where $l \in \{1, 2, \ldots, L\}$. Then the sampling weight vector $\boldsymbol{\Omega}$ of $\boldsymbol{\mathcal{X}}^*$ is computed as:

$$\boldsymbol{\Omega}(l) = \mu_l \cdot \nu_l. \tag{11}$$

AGS uses $\boldsymbol{\Omega}$ as sampling weights to sample a minimal subset from $\boldsymbol{\mathcal{X}}^*$ by the Monte-Carlo method.

C. Efficient Residual Sorting Strategy

When the total number of generated hypotheses is large, the computational cost of residual sorting in both MultiGS [16] and ModeSamp [18] is expensive because these methods sort the residuals of all input data after generating a batch of new hypotheses. Therefore, the efficiency of residual sorting greatly affects the efficiency of both MultiGS and ModeSamp.

In this paper, instead of sorting the residuals of all input data, we propose to only sort some of the top ranked residuals, to improve the efficiency of residual sorting, since the computation of the correlation between two data points (by Eq. (3)) only uses the index permutation of the top w ranked residuals, where w is the window size, and thus it is not necessary to sort the residuals of all input data.

Let c denote the total number of the generated hypotheses so far and g be the corresponding number of the retained top sorted residuals, where c and g are initialized to 0. After generating a batch (whose size is b) of new hypotheses (the influence of b will be evaluated in Section IV-C), the proposed residual sorting strategy, called Efficient Residual Sorting (ERS), updates the residual sorting vector as follows:

First, for the residual vector $\mathbf{r}^i$ of the ith data point $\boldsymbol{x}_i$, ERS sorts the residuals (including the top g sorted residuals and b residuals of $\boldsymbol{x}_i$ with regard to b newly generated hypotheses) of $\mathbf{r}^i$ in a non-descending order to achieve the sorted residual vector $\tilde{\mathbf{r}}^i = [\tilde{r}^i_{\kappa^i_1},\ \tilde{r}^i_{\kappa^i_2}, \ldots, \tilde{r}^i_{\kappa^i_{g+b}}]$.

Second, ERS updates g as:

$$g = \min(c, w + b). \tag{12}$$

where $c = c + b$ and $w = [\alpha * c]$, i.e., ERS retains all the elements of $\tilde{\mathbf{r}}^i$ when the first b hypotheses are generated; Otherwise, ERS only retains the top $w + b$ ranked residuals of $\tilde{\mathbf{r}}^i$ instead of the total c elements of $\tilde{\mathbf{r}}^i$.

Third, ERS first retains the top g elements of $\tilde{\mathbf{r}}^i$ (denoted as $\tilde{\mathbf{r}}^i = [\tilde{r}^i_{\kappa^i_1},\ \tilde{r}^i_{\kappa^i_2}, \ldots, \tilde{r}^i_{\kappa^i_g}]$). Then, ERS obtains the corresponding permutation vector $\boldsymbol{\kappa}^i = [\kappa^i_1,\ \kappa^i_2, \ldots, \kappa^i_g]$. Finally, let $\mathbf{r}^i = \tilde{\mathbf{r}}^i$. The efficiency of the proposed ERS will be evaluated in Section IV-B.

D. The Complete Method

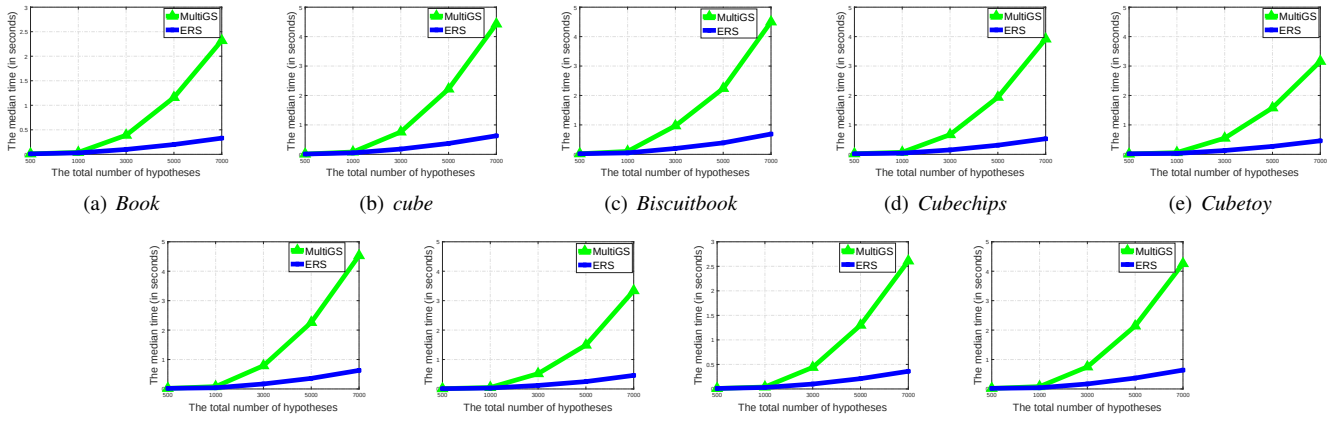
After introducing all the components of the proposed AGS, we present the complete AGS method, summarized in Algorithm 1. AGS contains two parameters, i.e., the batch size b and the ratio α of window size w. The batch size b controls the frequency of residual sorting. When we assign

Algorithm 1 The proposed AGS method

Input: data $\boldsymbol{\mathcal{X}}$, maximum number of samples M, ratio α and batch size b.

Output: a set of generated hypotheses $\boldsymbol{\Theta}$.

```
Initialization: Θ = null.
for c = 1 to M do
    if c ≤ b then
        Sample a minimal subset S_c by random sampling.
    else
        Sample a minimal subset S_c by the guided sampling strategy proposed in Section III-B.
    end if
    Estimate a hypothesis θ_c using the sampled minimal subset S_c.
    Append the absolute residual between x_i and θ_c to r^i for each x_i ∈ X.
    if c ≥ b and mod (c, b) = 0 then
        Update the window size w = [α * c].
        Sort r^i to obtain κ^i (Section III-C).
    end if
    Let Θ = Θ ∪ {θ_c}.
end for
```



(a) *Book* (b) *cube* (c) *Biscuitbook* (d) *Cubechips* (e) *Cubetoy*

(f) *Gamebiscuit* (g) *Breadcubechips* (h) *Toycubecar* (i) *Cubebreadtoychips*

Fig. 2. Comparisons on the efficiency of the proposed residual sorting strategy and the residual sorting strategy used in MultiGS on the nine image pairs. The total number of sampled hypotheses is shown in the x-axis, and the corresponding median time used for residual sorting is shown in the y-axis.

a small value to b, residuals will be sorted frequently, and the newly obtained hypotheses will be used to guide the sampling process timely. However, this is computationally expensive. In contrast, if the value of b is too large, the proposed method may not effectively use the newly obtained hypotheses to guide the sampling process, even though the computational cost is relatively lower. The parameter α controls the size of window w, which affects the efficiency of the correlation computation and the discriminative power of the computed correlations. The influence of these two parameters on the performance of the proposed AGS method will be evaluated in Section IV-C.

IV. Experiments

In this section, we first describe the datasets and performance measures used in the experiments in Section IV-A. We assess the efficiency of the proposed residual sorting strategy in Section IV-B. Then, we investigate the influence of the parameters of the proposed AGS, keypoint matching scores and information theoretic principles on the performance of AGS in Sections IV-C and IV-D, respectively. And then we evaluate the performance of the proposed AGS for data with various outlier rates and degeneracies in Sections IV-E and IV-F, respectively. Finally, we compare the proposed AGS with several state-of-the-art sampling methods in practical applications such as homography matrix and fundamental matrix estimation in Sections IV-G and IV-H, respectively.

We compare the proposed AGS with several sampling methods including MultiGS [16] and the sampling methods originally proposed in the following robust model fitting methods: PROSAC [13], LO-RANSAC [23], Proximity [24] and HMSS [28]. For convenience, these sampling methods are called the same as the corresponding robust fitting methods, respectively. The source codes of LO-RANSAC, PROSAC and MultiGS are provided by the authors of [16]. The source codes of Proximity and HMSS are from the authors of [19], [28], respectively. All experiments are run on a Windows 7 system with a 3.2 GHz i5 CPU.

A. Dataset and Performance Measures

The image pairs used in the following experiments are from the AdelaideRMF dataset [17]. This dataset provides inlier and outlier labels for all the image pairs and the corresponding keypoint matching scores.

For numerical stability, we perform the global coordinate normalization [34] on the test data before sampling. A hypothesis is estimated after a minimal subset is sampled and the corresponding residuals of input data with regard to the hypothesis are derived. HMSS [28] performs model selection by using the method proposed in [28], whereas all other competing methods including the proposed AGS use J-Linkage [19] to conduct model selection. Only the largest τ clusters are chosen after model selection as [17], where τ is the ground truth number of the structures in data. This is because that HMSS is a relatively accurate and efficient method by embedding its sampling strategy into the "fit-and-remove" model selection strategy, and its sampling strategy is not naturally involved in two-stage fitting methods (e.g., J-Linkage). While all other methods (i.e., AGS, MultiGS, PROSAC, LO-RANSAC and Proximity) mainly concentrate on hypothesis generation, and J-Linkage is one of the efficient multi-structure model selection methods. For the performance measure, as in [28], we measure Clustering Accuracy (CA) by

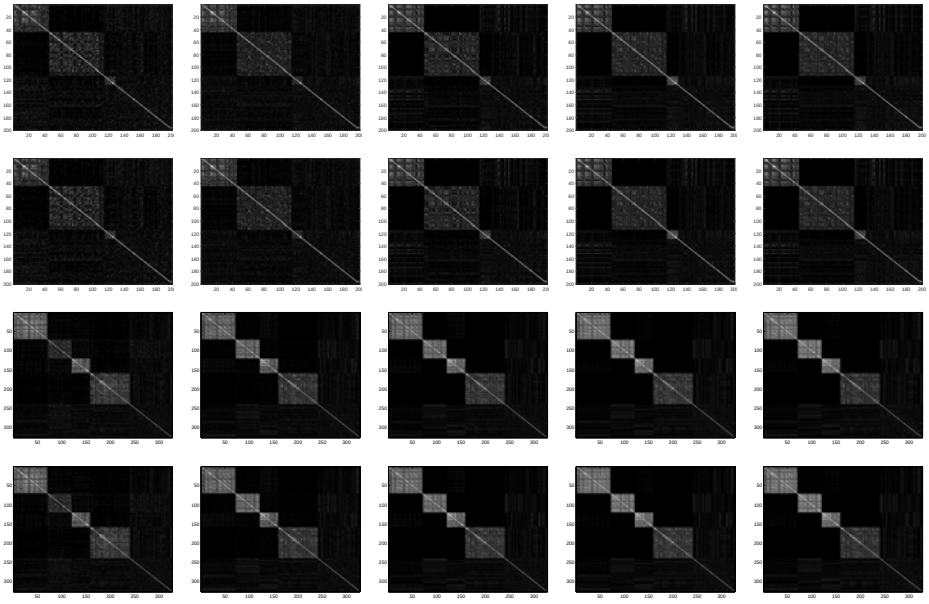
$$CA = \frac{number\ of\ correctly\ clustered\ points}{total\ number\ of\ points}. \quad (13)$$

In addition, we use the number of subsets and the number of clean subsets sampled within the given time to evaluate the performance of the competing methods as in [16], [17], [26].

B. Efficiency of the Proposed Residual Sorting Strategy

We demonstrate the efficiency of the proposed residual sorting strategy on the fundamental matrix estimation task. All the nine image pairs used in Table II are tested in the experiment. The 8-point algorithm [35] is used to estimate fundamental matrices from the sampled minimal subsets.

As shown in [16], MultiGS has achieved good performance when the batch size b and the ratio α of window size range

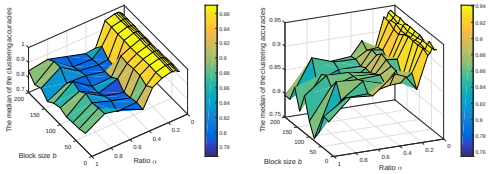


(a) $T = 500$ (b) $T = 1000$ (c) $T = 3000$ (d) $T = 5000$ (e) $T = 7000$

Fig. 3. The correlation matrices obtained by two different residual sorting strategies on the image pairs of *Toycubecar* (the first two rows) and *Cubebreadtoychips* (the last two rows). We use MultiGS (the first and the third rows) and ERS (the second and the fourth rows) as the residual sorting strategies. (a) to (e) The results obtained by MultiGS and ERS after generating $500, 1000, 3000, 5000$ and 7000 hypotheses, respectively. *Toycubecar* and *Cubebreadtoychips* have three and four model instances, respectively.

within [10.00, 100.00] and [0.05, 0.50], respectively. Since the proposed AGS adopts the framework of MultiGS [16], we compare the efficiency of the proposed residual sorting strategy ERS with that used in MultiGS. We set $b = 50$ and the window size $w = [0.05 * c]$, and generate 500, 1,000, 3,000, 5,000, and 7,000 hypotheses, respectively. The median time used for residual sorting over 20 runs is plotted in Fig. 2.

As shown in Fig. 2, the residual sorting strategy used in MultiGS is as efficient as the proposed method when the total number of the generated hypotheses is less than or equal to 1,000. However, when the total number of the generated hypotheses is larger than 3,000, the computational time of the residual sorting strategy used in MultiGS increases dramatically while that of ERS only increases slightly. Note that based on the same parameter settings, MultiGS generates about 1,500 hypotheses within 10 seconds for the *Cubebreadtoychips* image pair, while AGS generates about 6,500 hypotheses for that image pair. Thus, ERS can effectively improve the efficiency of AGS for hypotheses generation. Moreover, on the image pairs of *Toycubecar* and *Cubebreadtoychips*, we compare the correlation matrices computed from the permutation vector obtained by the residual sorting strategy used in MultiGS and the proposed ERS, respectively. *Toycubecar* and *Cubebreadtoychips* are the challenging image pairs, since the lowest inlier rates of the model instances in the two image pairs are 7% and 12%, respectively. The results in Fig. 3 show that correlation matrices obtained by ERS are as discriminative as those obtained by the sorting strategy of MultiGS for different numbers of hypotheses. Thus, the proposed ERS is an efficient and effective residual sorting strategy.



(a) *Barrsmith* (b) *Johnsona*

Fig. 4. The influence of the parameters b and α on the performance of AGS. The median of the clustering accuracies is obtained by AGS under different combinations of b and α on the image pairs of *Barrsmith* and *Johnsona*.

C. Influence of the Parameters b and α

On the homography matrix estimation task, we evaluate the influence of both the block size b and the ratio α of the window size w on the performance of AGS. We use the image pairs of *Barrsmith* and *Johnsona* for evaluation. The Direct Linear Transformation (DLT) [35] is used to estimate a homography matrix from a data subset. b and α are set within the ranges of [20.00, 200.00] and [0.02, 1.00], respectively. For each combination of b and α, AGS is performed 20 times, and each repetition is run 5 seconds. Fig. 4 shows the median of the clustering accuracies obtained by AGS over 20 runs.

From Fig. 4(a), we can observe that, the performance of AGS changes slightly when the value of b varies within [20.00, 200.00]. It can also be found that AGS achieves good performance when the value of α ranges between 0.02 and 0.20. However, the clustering accuracy obtained by AGS decreases dramatically as the value of α is larger than 0.20. This is because that, as α increases, the computational cost of Eq. (3) significantly increases, leading to the sharp drop of the

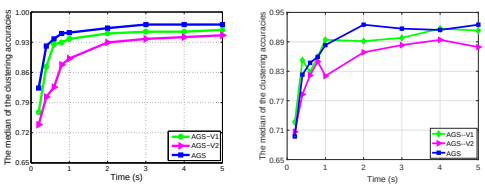
total number of samples in the given time. Even worse, the discriminative power of the correlations is also reduced, and thus the probability of sampling a clean solution decreases. In addition, α must be greater than 0 because the window size w is equal or greater than 1 (see Section III-A). Thus, we do not test the case when $\alpha = 0$. A similar conclusion for the *Johnsona* image pair can be drawn from Fig. 4(b). Based on the experimental results, we fix $b = 100$ and $\alpha = 0.1$ for AGS to achieve good performance in all the following experiments.

D. Influence of Keypoint Matching Scores and Information Theoretic Principles

In this section, we examine the influence of keypoint matching scores and information theoretic principles on the performance of the proposed AGS on the image pairs of *Barrsmith* and *Johnsona*. We compare the performance of the proposed method, the proposed method without using matching scores (called AGS-V1), and the proposed method without using both information theoretic principles and matching scores (called AGS-V2). Note that, we set the same score (i.e., 1.0) to each data point for AGS-V1 and AGS-V2. Fig. 5 shows the median of the clustering accuracies obtained by the three methods over 20 repetitions.

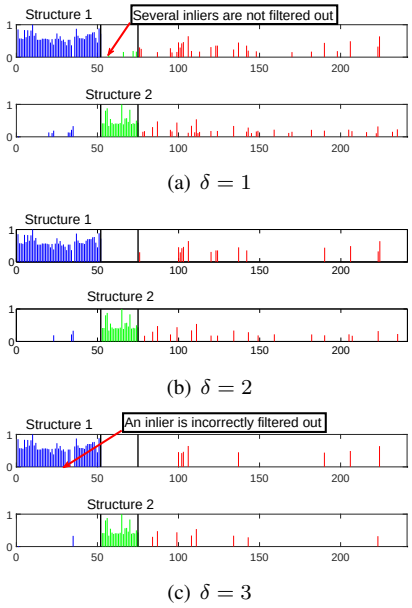
From Fig. 5(a), we find that both AGS and AGS-V1 can achieve relatively good performance. However, as the time budget increases, AGS-V1 is slower than AGS in increasing the median of the clustering accuracies. For example, the median of the clustering accuracies obtained by AGS increases to 0.95 within 0.80 second, while AGS-V1 needs to spend about 1.50 seconds on achieving the same accuracy. In addition, we find that the performance of AGS-V2 is worse than those of both AGS-V1 and AGS. A similar conclusion can be obtained for the *Johnsona* image pair from Fig. 5(b). As a result, both the matching scores and information theoretic principles used in AGS are beneficial to guide sampling.

In addition, we evaluate the influence of the parameter δ in Eq. (5) on the performance of the proposed method. The insight behind δ is that, the larger the value of δ is, the more data with low correlation values are filtered out. In this experiment, we filter out the data with low correlation values shown in Fig. 1(b) by evaluating different values of δ. The results are shown in Fig. 6. When $\delta = 1$ (as shown in the first row of Fig. 6(a)), several inliers of the second structure are not correctly filtered out. When $\delta = 2$ (as shown in the first row of Fig. 6(b)), all inliers of the second structure are correctly filtered out. When $\delta = 3$ (as shown in the first row of Fig. 6(c)), although more outliers are correctly filtered out than that in Fig. 6(b), an inlier of the first structure is incorrectly filtered out. A similar conclusion for the second structure can be obtained from the second row of Fig. 6(a) to (c). Therefore, to remain as many inliers of the current structure as possible, and filter out as many gross outliers and pseudo-outliers as possible, δ is set to be 2 in all the following experiments.



(a) *Barrsmith* (b) *Johnsona*

Fig. 5. The influence of both keypoint matching scores and information theoretic principles on the performance of AGS. The performance of the proposed method (i.e., AGS), of that without using matching scores (i.e., AGS-V1) and of that without using both information theoretic principles and matching scores (i.e., AGS-V2) are shown in different colors.



(a) $\delta = 1$

(b) $\delta = 2$

(c) $\delta = 3$

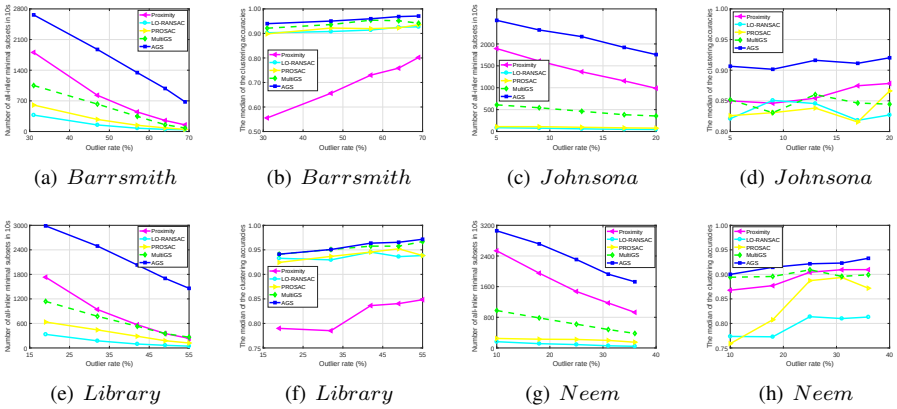
Fig. 6. The influence of the parameter δ on the performance of AGS. In the first (second) row of (a), (b) and (c), the data with lower correlation values are filtered out, and the remaining data are used to sample minimal subsets for the first (second) structure.

E. Performance under Various Outlier Rates

In this section, we evaluate the performance of the proposed AGS under different outlier rates on the tasks of homography matrix estimation and fundamental matrix estimation. The median number of all-inlier (i.e., clean) minimal subsets and the median of the clustering accuracies over 20 runs are shown in Fig. 7 and Fig. 8, respectively.

Homography matrix estimation. The experiments are evaluated on the image pairs of *barrsmith*, *Johnsona*, *Library* and *Neem*. In each run, we use a subset of the data, which contains all the inliers and L outliers. L outliers are randomly selected from all the outliers. For example, the image pair *barrsmith* includes 75 inliers and 166 outliers. The value of L gradually increases from 33 to 165 with an interval of 33. The outlier rate is computed by $100\% \cdot L/(L + 75)$.

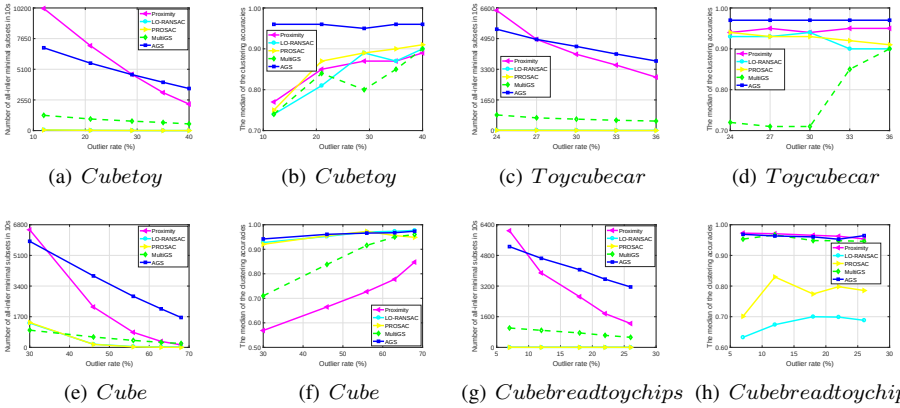
Although the total number of minimal subsets sampled by the proposed AGS is only about 70% of those obtained by Proximity, LO-RANSAC and PROSAC within 5 seconds, the total number of minimal subsets sampled by AGS is about 2 to 3 times of that obtained by MultiGS. This shows that, the proposed AGS significantly improves the computational efficiency compared with MultiGS.



(a) *Barrsmith* (b) *Barrsmith* (c) *Johnsona* (d) *Johnsona*

(e) *Library* (f) *Library* (g) *Neem* (h) *Neem*

Fig. 7. Performance for homography matrix estimation on the *barrsmith*, *Johnsona*, *Library* and *Neem* image pairs with various outlier rates. (a), (c), (e) and (g): The median number of all-inlier minimal subsets sampled in 5 seconds. (b), (d), (f) and (h): The median of the clustering accuracies.



(a) *Cubetoy* (b) *Cubetoy* (c) *Toycubecar* (d) *Toycubecar*

(e) *Cube* (f) *Cube* (g) *Cubebreadtoychips* (h) *Cubebreadtoychips*

Fig. 8. Performance for fundamental matrix estimation on the *Cubetoy*, *Toycubecar*, *Cube* and *Cubebreadtoychips* image pairs with various outlier rates. (a), (c), (e) and (g): The median number of all-inlier minimal subsets sampled in 10 seconds. (b), (d), (f) and (h): The median of the clustering accuracies.

From Fig. 7(a), (c), (e) and (g), we can see that the total numbers of clean minimal subsets sampled by all competing methods decrease as the outlier rate increases. The total number of clean minimal subsets sampled by AGS is about 2.5 times of that obtained by MultiGS. The results show that the proposed AGS significantly improves the sampling efficiency compared with MultiGS. Moreover, AGS can also sample more clean minimal subsets than Proximity, LO-RANSAC and PROSAC. Fig. 7(b), (d), (f) and (h) show that AGS achieves the highest clustering accuracy compared with all the other four competing methods. Especially on the image pair *Johnsona*, the gap between AGS and the best of all the other four competing methods is large. This benefits from the effectiveness of the proposed guided sampling strategy and the efficiency of the proposed residual sorting strategy in AGS.

Fundamental matrix estimation. The experiments are evaluated on the image pairs of *Cubetoy*, *Toycubecar*, *Cube* and *Cubebreadtoychips*. The experiments follow the similar experimental settings as those used in homography matrix estimation.

The total number of minimal subsets sampled by the proposed AGS is only about 1/2 of those obtained by Proximity, LO-RANSAC and PROSAC within 10 seconds. However, the total number of minimal subsets sampled by AGS is about 3.5 times of that obtained by MultiGS. This shows that the proposed AGS is much more computationally efficient than MultiGS.

Fig. 8 shows that the proposed AGS can usually sample more clean minimal subsets than all the other four competing sampling methods. On the image pairs of *Cubetoy* and *Toycubecar*, AGS achieves the highest clustering accuracy compared with all the four other competing methods. On the two other image pairs (i.e., *Cube* and *Cubebreadtoychips*), AGS achieves the highest or the second highest clustering accuracy. Although LO-RANSAC occasionally achieves the highest clustering accuracy for the single-structure data *Cube*, it achieves the lowest clustering accuracy for the four-structure data *Cubebreadtoychips* due to the high outlier rate in this image pair. In contrast, Proximity usually achieves the highest clustering accuracy for the *Cubebreadtoychips* image pair, while it achieves the lowest clustering accuracy for the *Cube* image pair because that it uses the spatial constraint. MultiGS achieves relatively good performance on the two image pairs, but its performance is not as good as that of AGS.

In summary, the proposed AGS can generate more accurate hypotheses, and it usually samples more number of clean minimal subsets than all the other competing methods within the same time budget.

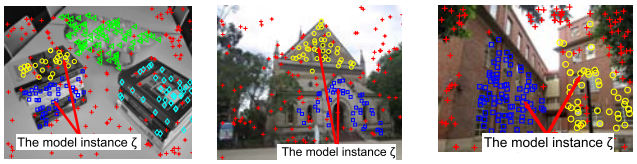
F. Performance under Degeneracies

In this section, we evaluate the performance of the competing sampling methods under degeneracies on the image pairs of *Dinobooks*, *Elderhalla* and *Sene* (see Fig. 9), following the similar experimental settings as the recent work [26]. For the 8-point estimation algorithm, a degenerate fundamental matrix is achieved when more than six of eight matches in a minimal subset are selected from the same plane [26]. This often occurs in the scene where there exists a dominant plane [36], [37].

Each of the image pairs of *Dinobooks*, *Elderhalla* and *Sene* contains one model instance (called the model instance ζ) of fundamental matrix, which includes one "dominant plane" D and one "off-plane" O. In the experiments, data subsets are created by choosing all inliers of D and part of the inliers of O. The ratio β between the inliers of the D plane and the inliers of both the D and O planes is computed as [26]:

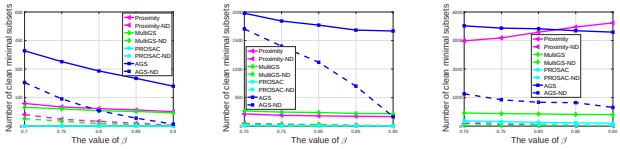
$$\beta = n_D/(n_O + n_D), \tag{14}$$

where n_O and n_D are the numbers of the inliers chosen from the D and O planes, respectively. We increase the value of β from 0.70 to 0.90 with an interval of 0.05 by gradually increasing the randomly selected inliers from O. For each value of β, 100 data subsets are created. Each method is limited to sample 10 seconds on each created data subset. The total numbers of both clean minimal subsets and non-degenerate clean minimal subsets sampled by the four sampling methods for the model instance ζ are recorded. Note that the performance of the other competing methods is not evaluated. This is because that both LO-RANSAC and HMSS may sample larger-than-minimal subsets. In addition, the inliers of the other two model instances are retained for the *Dinobooks* image pair, which case is more challenging than that in [26]. We show the median of the results over the 100 created data subsets with various values of β in Fig. 10.



(a) Dinobooks (b) Elderhalla (c) Sene

Fig. 9. The model instance ζ in the used image pairs (only showing one of the two views) contains two planes: the "dominant plane" is marked by the *blue squares*, and the "off-plane" is marked by the *yellow circles*.



(a) *Dinobooks* (b) *Elderhalla* (c) *Sene*

Fig. 10. Performance comparisons of different guided sampling methods under degeneracies. The numbers of both clean minimal subsets (solid lines) and *non-degenerate* clean minimal subsets (dashed lines) sampled for the model instance ζ by different sampling methods with different values of β.

As shown in Fig. 10(a), the numbers of both clean minimal subsets and non-degenerate clean minimal subsets sampled by AGS are significantly more (about 3 times more) than those obtained by Proximity and MultiGS, respectively, when $\beta \leq 0.85$. Even in the extreme case (i.e., $\beta = 0.90$, where only five inliers are selected from the "off-plane" O), AGS also outperforms Proximity and MultiGS in term of the numbers of both clean and non-degenerate clean minimal subsets sampled. PROSAC cannot sample any one clean minimal subset because *Dinobooks* has a high outlier rate (including pseudo and gross outliers). A similar conclusion for the *Elderhalla* image pair can be drawn from Fig. 10(b).

From Fig. 10(c), we can see that, the number of clean minimal subsets sampled by Proximity is comparable with that of AGS, and it is the most or the second most due to more number of inliers and the lower outlier rate in the $Sene$ image pair. However, AGS samples the highest number of non-degenerate clean minimal subsets compared with all the other three competing methods, and the number of the non-degenerate clean minimal subsets sampled by AGS is 10 times more than that obtained by Proximity. PROSAC and MultiGS can sample clean and non-degenerate clean minimal subsets. Nevertheless, the numbers of clean and non-degenerate clean minimal subsets obtained by the two methods are about only 10% of that obtained by AGS.

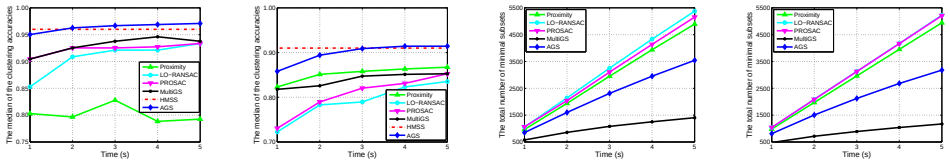
G. Comparisons on Homography Matrix Estimation

We investigate the evolution of both the clustering accuracies and the total number of minimal subsets obtained by different methods versus running time within 5 seconds. The obtained results are given in Fig. 11.

The evolution of the median of the clustering accuracies as increasing the sampling time is shown in Fig. 11(a) and (b). We note that HMSS finished the fitting process for each image pair in less than 1 second. For comparison, we report the same median clustering accuracies within 5 seconds for HMSS. From the figures, we find that the proposed AGS is the most effective method in increasing the median of the clustering accuracies versus running time. Especially, on the *Barrsmith* image pair, AGS achieves the highest median value of the clustering accuracies within 2 seconds compared with all the other competing methods. In other words, AGS is more than 2.5 times faster than all the other competing methods in generating promising hypotheses.

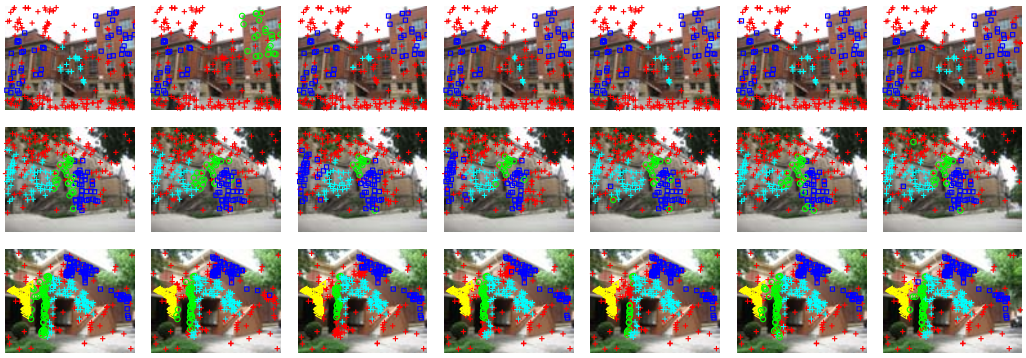
The evolution of the total number of minimal subsets as increasing the sampling time is shown in Fig. 11(c) and (d). We do not show the results for HMSS, because it finished the fitting process within less than 1 second. Fig. 11(c) and (d) show that, the total number of minimal subsets sampled by AGS is only about 3/5 of those obtained by Proximity, LO-RANSAC and PROSAC in 5 seconds. However, the total number of minimal subsets sampled by AGS is about 2.5 times of that obtained by MultiGS in 5 seconds.

To provide qualitative comparisons on the performance of these methods, the clustering results on three image pairs are shown in Fig. 12. In addition, the quantitative results obtained by the six competing methods over 100 runs are reported in



(a) *Barrsmith* (b) *Johnsona* (c) *Barrsmith* (d) *Johnsona*

Fig. 11. Evolution of both the clustering accuracies and the total number of minimal subsets obtained by different methods versus running time within 5 seconds. The median of the clustering accuracies obtained by the six methods on (a) *Barrsmith* and (b) *Johnsona*, respectively. The total number of minimal subsets sampled by different methods on (c) *Barrsmith* and (d) *Johnsona*, respectively.



(a) (b) (c) (d) (e) (f) (g)

Fig. 12. Qualitative comparisons on homography detection (only showing one of the two views). (a) shows the ground truth whereas (b) to (g) show the segmentation results obtained by Proximity, LO-RANSAC, PROSAC, MultiGS, HMSS and the proposed AGS, respectively. The *Barrsmith*, *Elderhallb* and *Johnsona* image pairs are shown in the top-down order.

Table I. From the results, we can see that, the proposed AGS achieves the highest median of the clustering accuracies on all test image pairs except for the *Napiera* image pair. On this image pair, the median of the clustering accuracies obtained by AGS is only slightly lower than the best result obtained by PROSAC. On eight of the nine image pairs, AGS achieves the lowest standard variance of the clustering accuracies. Moreover, we emphasize that the clustering accuracy obtained by AGS is better than (or equal to) that obtained by MultiGS on all image pairs.

In addition to clustering accuracy, we also report the total number of subsets and the number of all-inlier subsets sampled by all the competing methods within the given time. As stated before, the total number of subsets sampled by the proposed AGS is less than those sampled by Proximity, LO-RANSAC and PROSAC, but it is more than that sampled by MultiGS. However, on all image pairs except for the *Oldclassicswing* image pair, the total number of all-inlier subsets sampled by AGS is much more than those sampled by the other five competing methods. Especially, AGS is about 3 to 8 times more than MultiGS in generating all-inlier minimal subsets.

HMSS and MultiGS obtain fairly good performance for all image pairs. Proximity, LO-RANSAC and PROSAC achieve good performance on most test image pairs as well. However, due to the assumption of that the inliers from the same structure are spatially close, the results obtained by Proximity are bad if a structure is separated by another structure, such as the *Barrsmith* image pair (see Fig. 12). LO-RANSAC and PROSAC perform badly on highly contaminated data such as the *Elderhallb* image pair due to the influence of outliers.

We note that Table I shows that LO-RANSAC and HMSS cannot sample any all-inlier subset for the second model instances of the image pairs *Barrsmith* and *Elderhallb*. However, Fig. 12 shows that LO-RANSAC and HMSS can find some inliers for the second model instances of the image pairs *Barrsmith* and *Elderhallb*. This is because that: (1) The first row of Fig. 13(b) shows that a cross-structure subset is sampled by LO-RANSAC on the *Barrsmith* image pair. The cross-structure subset contains the inliers of both the first and the second model instances, and one gross outlier. Based on the cross-structure subset, J-Linkage finds the inliers of the second model instance shown in the first row of Fig. 13(c). However, the hypothesis generated from a cross-structure subset is inaccurate. This is consistent with the conclusion that the

(a) (b) (c)

Fig. 13. Some results obtained by LO-RANSAC and HMSS on the *Barrsmith* and *Elderhallb* image pairs, respectively, shown in the top-down order: (a) the ground truth, (b) the "Clean" samples and (c) the corresponding results.

TABLE I

THE PERFORMANCE OBTAINED BY THE SIX SAMPLING METHODS OVER 100 REPETITIONS FOR HOMOGRAPHY MATRIX ESTIMATION. LEGEND: #TOTAL REPORTS THE TOTAL NUMBER OF SUBSETS SAMPLED BY THE SIX METHODS WITHIN 5 SECONDS; #S-i (THE NUMBER OF INLIERS BELONGING TO THE iTH STRUCTURE, THE INLIER RATIO OF THE iTH STRUCTURE) REPORTS THE TOTAL NUMBER OF ALL-INLIER SUBSETS SAMPLED BY DIFFERENT METHODS FOR THE iTH STRUCTURE; MEDIAN CA AND STD CA REPORT THE MEDIAN AND THE STANDARD VARIANCE OF THE OBTAINED CLUSTERING ACCURACIES OVER 100 REPETITIONS, RESPECTIVELY. THE BEST RESULT OBTAINED BY THE SIX METHODS ON EACH MEASURE CRITERIA IS BOLDFACED.

Data		Proximity	LO-RANSAC	PROSAC	MultiGS	HMSS	AGS
Bonython	#Total	5359	5573	**5708**	1494	191	3706
	#S-1 (52, 26%)	275	58	104	226	25	**946**
	Median CA	0.99	0.99	**1.00**	**1.00**	0.97	**1.00**
	Std CA	0.01	0.01	**0.00**	0.01	0.01	**0.00**
Unionhouse	#Total	5113	5414	**5433**	1186	230	3240
	#S-1 (78, 23%)	84	46	90	111	28	**906**
	Median CA	0.95	**0.98**	**0.98**	**0.98**	0.97	**0.98**
	Std CA	0.03	**0.00**	0.01	0.01	**0.00**	**0.00**
Barrsmith	#Total	5331	**5603**	5573	1439	362	3535
	#S-1 (52, 22%)	110	34	43	73	7	**532**
	#S-2 (23, 10%)	39	0	1	9	2	**83**
	Median CA	0.79	0.93	0.93	0.95	0.96	**0.97**
	Std CA	0.05	0.03	0.04	0.04	0.06	**0.01**
Library	#Total	5405	5686	**5750**	1493	383	3768
	#S-1 (50, 23%)	152	25	64	147	10	**768**
	#S-2 (46, 21%)	65	15	47	112	2	**636**
	Median CA	0.86	0.93	0.94	0.96	0.96	**0.98**
	Std CA	0.06	0.04	0.03	0.03	**0.00**	0.01
Napiera	#Total	5228	5484	**5587**	1265	450	3381
	#S-1 (30, 10%)	3	0	2	3	0	**45**
	#S-2 (82, 27%)	154	50	119	105	8	**517**
	Median CA	0.83	0.95	**0.96**	0.94	0.89	0.95
	Std CA	0.05	0.03	**0.02**	0.04	**0.02**	**0.02**
Oldclassicswing	#Total	5183	5418	**5480**	1142	219	3167
	#S-1 (185, 49%)	**1819**	325	1448	422	83	1486
	#S-2 (71, 19%)	167	6	13	91	2	**485**
	Median CA	**0.97**	0.94	0.96	**0.97**	0.96	**0.97**
	Std CA	0.02	0.03	0.02	0.02	**0.00**	**0.00**
Elderhallb	#Total	5316	5603	**5710**	1363	382	3580
	#S-1 (42, 16%)	160	4	14	44	2	**182**
	#S-2 (28, 11%)	15	1	3	5	0	**47**
	#S-3 (63, 25%)	279	29	87	145	9	**815**
	Median CA	0.93	0.77	0.77	0.93	**0.94**	**0.94**
	Std CA	0.02	0.07	0.05	0.07	0.02	**0.01**
Neem	#Total	5298	5539	**5734**	1411	344	3631
	#S-1 (64, 27%)	608	29	95	222	19	**803**
	#S-2 (43, 18%)	177	5	27	58	3	**450**
	#S-3 (46, 19%)	88	6	20	131	1	**360**
	Median CA	0.91	0.80	0.85	0.91	0.92	**0.93**
	Std CA	**0.02**	0.06	0.04	0.03	**0.02**	**0.02**
Johnsona	#Total	5143	5346	**5449**	1151	425	3181
	#S-1 (77, 21%)	312	10	16	116	5	**522**
	#S-2 (91, 24%)	314	21	45	147	21	**562**
	#S-3 (69, 19%)	88	6	5	35	2	**152**
	#S-4 (58, 16%)	241	3	7	67	13	**421**
	Median CA	0.88	0.83	0.85	0.85	**0.92**	**0.92**
	Std CA	**0.02**	0.06	0.04	0.06	0.04	**0.02**

success of the hypothesize-and-verify method relies on hitting at least one all-inlier minimal subset [16]. Moreover, we show later that the cross-structure subsets may lead to overfitting on the high order geometric model (such as fundamental matrix). (2) The second row of Fig. 13(b) shows that a cross-structure subset is sampled by HMSS on the *Elderhallb* image pair. The cross-structure subset contains the inliers of both the first and the second model instances. Due to the fact that the first and the second model instances are in contact, the hypothesis generated from the cross-structure subset is relatively accurate.

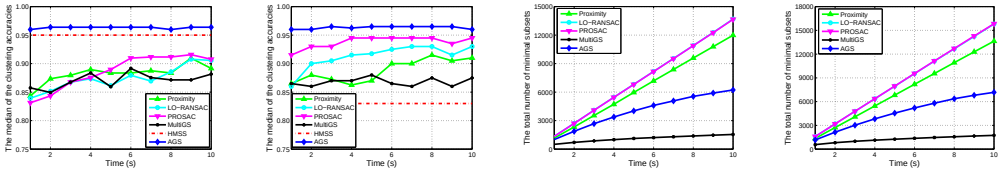
H. Comparisons on Fundamental Matrix Estimation

Following the similar experimental setup as in Section IV-G, Fig. 14 shows the the evolution of both the clustering accuracies and the total number of minimal subsets obtained by different methods. The evolution of the median of the clustering accuracies as increasing the sampling time is shown in Fig. 14(a) and (b). We observe that the proposed AGS is the most effective method to increase the median of the clustering accuracies. Moreover, AGS achieves the highest median of the clustering accuracies in less than 1 second compared with all the other competing methods. In other words, AGS is more than 10 times faster than all the other competing methods in generating promising hypotheses. The evolution of the total number of minimal subsets is shown in Fig. 14(c) and (d). A similar conclusion can be drawn as that obtained from Fig. 11(c) and (d).

To qualitatively compare the performance of the six competing methods, the clustering results on three image pairs are shown in Fig. 15. Moreover, the quantitative results obtained by the six methods over 100 runs are shown in Table II. From the results, we find that the proposed AGS obtains the highest median of the clustering accuracies on all test image pairs except for two image pairs *GameBiscuit* and *Cubebreadtoychips*. However, on the two image pairs, the median of the clustering accuracies obtained by AGS are still comparable with the best results. Moreover, on all test image pairs, the median of the clustering accuracies achieved by AGS is better than (or equal to) those achieved by MultiGS.

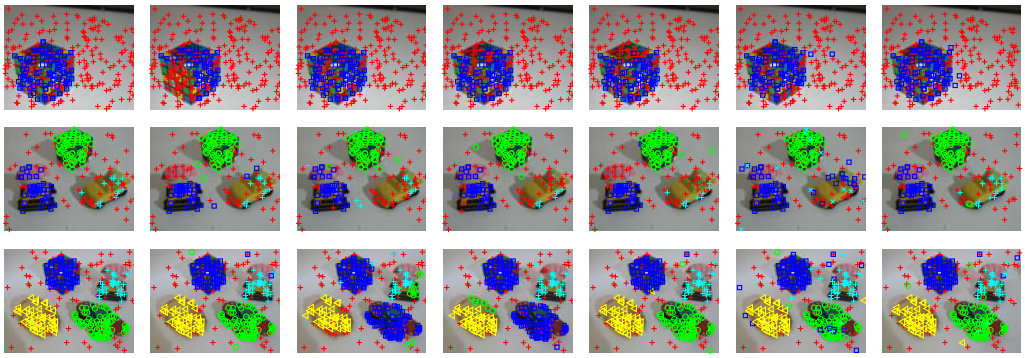
The total number of minimal subsets sampled by the proposed AGS is less than those obtained by Proximity, LO-RANSAC and PROSAC, but it is more than that obtained by MultiGS. However, on all test image pairs, the numbers of all-inlier subsets sampled by AGS are much more than those sampled by all the other competing methods. Especially, the number of all-inlier minimal subsets sampled by AGS is 5 to 7 times more than that obtained by MultiGS.

MultiGS also achieves fairly good solutions for all test image pairs. However, because its hypothesis generation process is time-consuming, it is incapable of sampling accurate hy-



(a) *Cubetoy* (b) *Toycubecar* (c) *Cubetoy* (d) *Toycubecar*

Fig. 14. Evolution of both the clustering accuracies and the total number of minimal subsets obtained by different methods. The median of the clustering accuracies obtained by the six competing methods on (a) *Cubetoy* and (b) *Toycubecar*, respectively. The total number of minimal subsets sampled by the six competing methods on (c) *Cubetoy* and (d) *Toycubecar*, respectively.



(a) (b) (c) (d) (e) (f) (g)

Fig. 15. Qualitative comparisons on two-view motion segmentation (only showing one of the two views). (a) shows the the ground truth whereas (b) to (g) show the segmentation results achieved by Proximity, LO-RANSAC, PROSAC, MultiGS, HMSS and the proposed AGS, respectively. The *Cube*, *Toycubecar* and *Cubebreadtoychips* image pairs are shown in the top-down order.

potheses for the structures with low inlier rates. For example, the median of the clustering accuracies obtained by MultiGS is low on the *Toycubecar* image pair because it cannot sample an accurate hypothesis for the structures of *toy* and *car* (marked in the blue squares and the cyan circles) within the given time (see Fig. 15 (e)).

HMSS, LO-RANSAC and PROSAC obtain high median clustering accuracies for single-structure image pairs, but they achieve low median clustering accuracies for most multi-structure image pairs (e.g., *Breadcubechips* and *Cubebreadtoychips*) due to the high outlier rates caused by both gross outliers and pseudo-outliers. In addition, because of the use of the "fit-and-remove" framework, HMSS may wrongly estimate the model parameters for multi-structure image pairs (e.g., *Toycubecar*–see Fig. 15 (f)). LO-RANSAC and PROSAC may over-fit on multi-structure image pairs (e.g., *Cubebreadtoychips*), because they only consider the single-structure case and try to maximize the consensus set, resulting in sampling cross-structure minimal subsets (see Fig. 15 (c) and (d)).

Proximity assumes that the inliers from the same model instance are spatially close, and it samples minimal subsets by using local constraints. Therefore, Proximity cannot sample an accurate hypothesis for an image pair, which only contains one model instance and the model instance occupies a large proportion of the area of the image. This is because that the spans of the minimal subsets sampled by Proximity are small and the hypotheses generated from these minimal subsets are inaccurate [26]. For example, Fig. 15 (b) shows that many inliers of the *Cube* image pair are not detected.

V. Conclusions

In this paper, a new hypothesis generation method, called AGS, is proposed. AGS combines the merits of both residual sorting and keypoint matching scores to effectively generate accurate hypotheses for robust model fitting methods. Moreover, a new residual sorting strategy, which reduces the computational cost of residual sorting in AGS, is proposed to increase the efficiency of AGS. Experimental results on a number of challenging real image pairs show that, the proposed AGS outperforms several state-of-the-art sampling methods, and it achieves the best results on most of the test data. Especially on some challenging data, AGS is several times or even one order faster than the other competing sampling methods in generating promising hypotheses.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grants U1605252, 61702101, 61472334, 61571379, 61701117, by the National Key Research and Development Plan (No. 2016YFC0801002), and by the China Postdoctoral Science Foundation 2017M622038. Tat-Jun Chin acknowledged funding under ARC DP160103490.

References

[1] X. Qian and C. Ye, "NCC-RANSAC: A fast plane extraction method for 3-D range data segmentation," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2771–2783, 2014.

TABLE II

THE PERFORMANCE OBTAINED BY THE SIX COMPETING SAMPLING METHODS OVER 100 RUNS FOR FUNDAMENTAL MATRIX ESTIMATION. THE SAME PERFORMANCE METRICS AS TABLE I ARE USED. THE BEST RESULT OBTAINED BY THE SIX METHODS ON EACH MEASURE CRITERIA IS BOLDFACED.

Data		Proximity	LO-RANSAC	PROSAC	MultiGS	HMSS	AGS
Book	#Total	13755	**17530**	17278	1819	92	7392
	#S-1 (105, 56%)	2918	193	235	588	20	**3904**
	Median CA	0.94	**0.99**	**0.99**	0.98	0.97	**0.99**
	Std CA	0.10	**0.01**	**0.01**	0.09	0.02	0.05
Cube	#Total	10931	**12219**	12110	1389	158	5521
	#S-1 (97, 32%)	124	16	2	225	8	**1721**
	Median CA	0.87	**0.97**	0.95	0.92	0.95	**0.97**
	Std CA	0.06	**0.02**	0.04	0.07	0.03	0.04
Biscuitbook	#Total	10089	**11021**	10969	1320	412	5067
	#S-1 (97, 28%)	213	10	1	223	33	**1387**
	#S-2 (82, 24%)	181	0	2	178	15	**1224**
	Median CA	0.97	0.96	0.96	0.90	0.93	**0.99**
	Std CA	0.06	0.04	0.04	0.07	**0.02**	0.05
GameBiscuit	#Total	10366	**11446**	11413	1319	371	5222
	#S-1 (73, 22%)	61	0	26	177	49	**1393**
	#S-2 (88, 27%)	542	2	0	208	0	**1279**
	Median CA	**0.93**	0.89	0.92	0.89	0.91	0.89
	Std CA	0.05	0.06	0.05	0.03	**0.02**	0.04
Cubechips	#Total	11467	**11682**	11649	1461	403	5551
	#S-1 (84, 30%)	579	12	6	245	10	**1618**
	#S-2 (57, 20%)	344	0	0	159	0	**1005**
	Median CA	**0.98**	**0.98**	**0.98**	**0.98**	0.93	**0.98**
	Std CA	**0.01**	0.02	0.04	0.02	0.05	**0.01**
Cubetoy	#Total	12312	**12983**	12915	1543	418	6065
	#S-1 (78, 31%)	925	3	2	293	26	**1718**
	#S-2 (72, 29%)	1175	2	1	292	17	**1682**
	Median CA	0.88	0.90	0.92	0.89	0.95	**0.96**
	Std CA	0.06	0.09	0.06	0.08	0.06	**0.03**
Breadcubechips	#Total	12261	**14325**	14253	1610	666	6395
	#S-1 (34, 15%)	146	0	0	80	0	**597**
	#S-2 (57, 25%)	681	0	1	237	12	**1557**
	#S-3 (58, 25%)	290	0	0	223	5	**1427**
	Median CA	**0.97**	0.72	0.72	**0.97**	0.88	**0.97**
	Std CA	0.02	0.04	0.07	0.05	0.18	**0.01**
Toycubecar	#Total	13572	16035	**16043**	1742	375	7202
	#S-1 (45, 23%)	848	0	0	179	2	**1293**
	#S-2 (69, 35%)	1754	13	6	373	35	**2496**
	#S-3 (14, 7%)	0	0	0	1	0	**24**
	Median CA	0.91	0.93	0.94	0.87	0.83	**0.96**
	Std CA	0.09	0.08	0.07	0.04	0.04	**0.02**
Cubebreadtoychips	#Total	9925	10759	**10780**	1351	975	5119
	#S-1 (71, 22%)	467	0	0	189	12	**1012**
	#S-2 (49, 15%)	184	0	0	103	4	**475**
	#S-3 (38, 12%)	46	0	0	60	0	**383**
	#S-4 (81, 25%)	319	0	1	215	10	**1183**
	Median CA	**0.97**	0.71	0.73	0.96	0.88	0.96
	Std CA	**0.03**	**0.03**	0.10	0.06	0.09	0.06

[2] M. Liang, H. Min, R. Luo, and J. Zhu, "Simultaneous recognition and modeling for learning 3-D object models from everyday scenes," *IEEE Trans. Cybern.*, vol. 45, no. 10, pp. 2237–2248, 2015.

[3] P. Kaur, K. J. Dana, F. A. Romero, and N. Gucunski, "Automated gpr rebar analysis for robotic bridge deck evaluation," *IEEE Trans. Cybern.*, vol. 46, no. 10, pp. 2265–2276, 2016.

[4] J. Chen, B. Jia, and K. Zhang, "Trifocal tensor-based adaptive visual trajectory tracking control of mobile robots," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3784–3798, 2017.

[5] L. Li, Q. Xu, V. Chandrasekhar, J.-H. Lim, C. Tan, and M. A. Mukawa, "A wearable virtual usher for vision-based cognitive indoor navigation," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 841–854, 2017.

[6] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, to be published, doi: 10.1007/s11263-018-1117-z.

[7] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM.*, vol. 24, no. 6, pp. 381–395, 1981.

[8] H. Isack and Y. Boykov, "Energy based multi-model fitting & matching for 3D reconstruction," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1146–1153.

[9] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J. Frahm, "USAC: A universal framework for random sample consensus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 2022–2038, 2013.

[10] R. Litman, S. Korman, A. Bronstein, and S. Avidan, "Inverting RANSAC: Global model detection via inlier rate estimation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5243–5251.

[11] 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, 2012.

[12] P. Meer, "Robust techniques for computer vision," *Emerging topics in computer vision*, pp. 107–190, 2004.

[13] O. Chum and J. Matas, "Matching with PROSAC-progressive sample consensus," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 220–226.

[14] 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, 2005.

[15] V. Fragoso, P. Sen, S. Rodriguez, and M. Turk, "EVSAC: Accelerating hypotheses generation by modeling matching scores with extreme value theory," in *Proceedings of IEEE International Conference on Computer Vision*, 2013, pp. 2472–2479.

[16] 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, 2012.

[17] H. S. Wong, T.-J. Chin, J. Yu, and D. Suter, "A simultaneous sample-and-filter strategy for robust multi-structure model fitting," *Computer Vision and Image Understanding*, vol. 117, no. 12, pp. 1755–1769, 2013.

[18] ——, "Mode seeking over permutations for rapid geometric model fitting," *Pattern Recognit.*, vol. 46, no. 1, pp. 257–271, 2013.

[19] R. Toldo and A. Fusiello, "Robust multiple structures estimation with J-linkage," in *Proceedings of European Conference on Computer Vision*, 2008, pp. 537–547.

[20] T. Lai, H. Wang, T. J. Chin, and W. L. Zhao, "Motion segmentation via a sparsity constraint," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 973–983, 2017.

[21] L. Magri and A. Fusiello, "T-Linkage: A continuous relaxation of J-Linkage for multi-model fitting," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3954–3961.

[22] D. Myatt, P. Torr, S. Nasuto, J. Bishop, and R. Craddock, "NAPSAC: high noise, high dimensional robust estimation-its in the bag," in *Proceedings of British Machine Vision Conference*, 2002, pp. 458–467.

[23] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," in *DAGM-Symp*, 2003, pp. 236–243.

[24] Y. Kanazawa and H. Kawakami, "Detection of planar regions with uncalibrated stereo using distributions of feature points." in *Proceedings of British Machine Vision Conference*, 2004, pp. 247–256.

[25] K. Ni, H. Jin, and F. Dellaert, "GroupSAC: Efficient consensus in

the presence of groupings," in *Proceedings of IEEE International Conference on Computer Vision*, 2009, pp. 2193–2200.
[26] Q. H. Tran, T.-J. Chin, W. Chojnacki, and D. Suter, "Sampling minimal subsets with large spans for robust estimation," *Int. J. Comput. Vis.*, vol. 106, no. 1, pp. 93–112, 2014.
[27] K. Lebeda, J. Matas, and O. Chum, "Fixing the locally optimized RANSAC," in *Proceedings of British Machine Vision Conference*, 2012, pp. 1–11.
[28] R. 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, 2016.
[29] T. Lai, H. Fujita, C. Yang, Q. Li, and R. Chen, "Robust model fitting based on greedy search and specified inlier threshold," *IEEE Trans. Ind. Electron.*, to be published, doi: 10.1109/TIE.2018.2881950.
[30] K. H. Lee and S. W. Lee, "Deterministic fitting of multiple structures using iterative MaxFS with inlier scale estimation," in *Proceedings of IEEE International Conference on Computer Vision*, 2013, pp. 41–48.
[31] M. Zuliani, C. S. Kenney, and B. Manjunath, "The MultiRANSAC algorithm and its application to detect planar homographies," in *Proceedings of IEEE Conference on Image Processing*, 2005, pp. 153–156.
[32] C. E. Shannon and W. Weaver, *A mathematical theory of communication*. University of Illinois Press, 1963.
[33] L. Ferraz, R. Felip, B. Martínez, and X. Binefa, "A density-based data reduction algorithm for robust estimators," *Pattern Recognition and Image Analysis*, pp. 355–362, 2007.
[34] R. I. Hartley, "In defense of the eight-point algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 6, pp. 580–593, 1997.
[35] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2004.
[36] O. Chum, T. Werner, and J. Matas, "Two-view geometry estimation unaffected by a dominant plane," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 772–779.
[37] J.-M. Frahm and M. Pollefeys, "RANSAC for (quasi-) degenerate data (QDEGSAC)," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 453–460.

Taotao Lai is currently a Postdoctoral Fellow in the College of Computer and Information Sciences at Fujian Agriculture and Forestry University, China. He received the Ph.D. degree in Computer Science and Technology from Xiamen University, China, in 2016. He has published several papers in the international journals including the IEEE TITS, PR, CVIU, etc. His research interests include robust model fitting and structure from motion.

Hanzi Wang is currently a Distinguished Professor of "Minjiang Scholars" in Fujian province and a Founding Director of the Center for Pattern Analysis and Machine Intelligence (CPAMI) at Xiamen University in China. He received his Ph.D degree in Computer Vision from Monash University. His research interests are concentrated on computer vision and pattern recognition including visual tracking, robust statistics, object detection, video segmentation, model fitting, optical flow calculation, 3D structure from motion, image segmentation and related fields. He was an Associate Editor for IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) from 2010 to 2015 and a Guest Editor of Pattern Recognition Letters (September 2009). He was the General Chair for ICIMCS2014, Program Chair for CVRS2012, Area Chair for ACCV2016, DICTA2010, Tutorial Chair for VALSE2017, Publicity Chair for ICIG2015 and IEEE NAS2012. He also served on the program committee (PC) of ICCV, ECCV, CVPR, ACCV, PAKDD, ICIG, ADMA, CISP, etc.

Yan Yan is currently an associate professor in the School of Information Science and Engineering at Xiamen University, China. He received the Ph.D. degree in Information and Communication Engineering from Tsinghua University, China, in 2009. He worked at Nokia Japan R&D center as a research engineer (2009-2010) and Panasonic Singapore Lab as a project leader (2011). He has published around 60 papers in the international journals and conferences including IEEE TPAMI, IJCV, IEEE TIP, IEEE TMM, IEEE TCYB, IEEE TITS, PR, KBS, ICCV, ECCV, ACM MM, ICPR, ICIP, etc. His research interests include computer vision and pattern recognition. He is a member of the IEEE.

Tat-Jun Chin received his BEng degree in mechatronics engineering from Universiti Teknologi Malaysia (UTM) in 2003, where he received the Vice Chancellors Commendation Award, and his PhD in computer systems engineering from Monash University in 2007, which was supported by the Endeavour Australia-Asia Award. He is currently an Associate Professor at The University of Adelaide. He is also an Associate Editor of the IPSJ Transactions on Computer Vision and Applications (TCVA) and Journal of Imaging (J. Imaging). Tat-Jun's research interest lies in optimisation for geometric vision, which covers areas such as 3D reconstruction, SLAM, and augmented reality. He has published more than 80 research articles on the subject, including an invited book submission. Tat-Jun has won significant awards for his research: a CVPR Award (2015), two DST Awards (2015 and 2017), and a BMVC Award (2018).

Jin Zheng received the B.S. degree in applied mathematics and informatics from the College of Science in 2001, and the M.S. degree from the School of Computer Science, Liaoning Technical University, Fuxing, China, in 2004, and the Ph.D. degree from the School of Computer Science and Engineering, Beihang University, Beijing, China, in 2009. She is currently an Associate Professor with Beihang University. Her current research interests include moving object detection and tracking, object recognition, image enhancement, video stabilization, and video mosaic.

Bo Li received the B.S. degree in computer science from Chongqing University in 1986, the M.S. degree in computer science from Xian Jiaotong University in 1989, and the Ph.D. degree in computer science from Beihang University in 1993. Now he is a professor of Computer Science and Engineering at Beihang University, the Director of Beijing Key Laboratory of Digital Media, and has published over 100 conference and journal papers in diversified research fields including digital video and image compression, video analysis and understanding, remote sensing image fusion and embedded digital image processor.