Accelerated Hypothesis Generation for Multi-Structure Data via Preference Analysis

Tat-Jun Chin, *Member, IEEE,* Jin Yu,and David Suter, *Senior Member, IEEE*

Abstract—Random hypothesis generation is integral to many robust geometric model fitting techniques. However it is also computationally expensive, especially for higher-order geometric models and heavily contaminated data. We propose a fundamentally new approach to accelerate hypothesis sampling by guiding it with information derived from residual sorting. We show that residual sorting innately encodes the probability of two points to have arisen from the same model; and is obtained without recourse to domain knowledge (*e.g.* keypoint matching scores) typically used in previous sampling enhancement methods. More crucially our approach encourages sampling within coherent structures and thus can rapidly generate all-inlier minimal subsets favoured by the robust criterion. Sampling within coherent structures also affords a natural ability to handle multi-structure data, a condition that is usually detrimental to other methods. Our sampling scheme offers substantial speed-ups on common computer vision tasks such as homography and fundamental matrix estimation. We show on many realistic data, especially those with multiple structures, that ours is the only method capable of retrieving satisfactory results within realistic time budgets. Matlab implementation is available from the authors' homepages.

Index Terms—Geometric model fitting, robust estimation, hypothesis generation, residual sorting, multiple structures

1 Introduction

Random hypothesis sampling is central to many state-of-the-art robust estimation techniques. The procedure is often embedded in the "hypothesise-and-verify" framework commonly found in methods such as Random Sample Consensus (RANSAC) [8] and Least Median Squares (LMedS) [20]. The goal of sampling is to generate many putative hypotheses of a given geometric model (*e.g.* the fundamental matrix), where each hypothesis is fitted on a randomly chosen *minimal subset* of the input data. The hypotheses are then scored in the verification stage according to a robust criterion (*e.g.* number of inliers, median of squared residuals).

The success of hypothesise-and-verify relies on "hitting" at least one minimal subset containing only inliers of a particular genuine instance of the geometric model. Let $\mathcal{X} = \{\boldsymbol{x}_i\}_{i=1}^N$ be a set of N input data. A minimal subset is a subset $\mathcal{S}$ of $\mathcal{X}$ of size p, where p is also the order of the geometric model, i.e. the number of parameters in the model. A minimal subset $\mathcal{S}$ which contains only inliers to a particular model instance of the model allows the determination the model parameters without being affected by outliers.

Under random hypothesis sampling [8], each member of $\mathcal{S}$ is chosen randomly without replacement. Assuming $N \gg p$ the probability that $\mathcal{S}$ contains all inliers is approximately $(1-\varepsilon)^p$, where $\varepsilon \in [0,1]$ is the *outlier contamination rate*, *i.e.*, the proportion of outliers in $\mathcal{X}$. Therefore on average $1/(1-\varepsilon)^p$ samples will be required before hitting an all-inlier minimal subset. Note that the probability value $(1-\varepsilon)^p$ decreases exponentially with p. In other words for large p's consecutively getting p inliers via pure random selection is non-trivial.

The analysis does not augur well for computer vision applications. Firstly, computer vision data is usually heavily contaminated with outliers due to imperfections in acquisition and preprocessing. Moreover in practical settings the data usually contains *multiple* instances of the geometric model or *structures* [24]. In such cases the effective outlier rate faced by a structure is contributed to by both the *gross* outliers and *pseudo*-outliers (*i.e.*, inliers of other valid structures). Secondly, many useful geometric models are of significant order in p. This is certainly true in multi-view geometry where robust estimators have seen widespread usage (*e.g.* $p = 4$ for homographies, $p = 7$ or 8 for fundamental matrices).

Fig. 1 shows an example of multi-structure data pertaining to the two-view motion segmentation problem. Point matches are established on the image pairs using SIFT keypoint matching [14], and the matches on a particular motion give rise to a fundamental matrix relation. By careful manual filtering, the matches are determined to consist of 113 outliers (40.50%) while the inliers are separated into 3 groups whose sizes are respectively 69 (24.73%), 68 (24.37%) and 29 (10.39%). Referring to the first structure and using the 8-point algorithm for fundamental matrix estimation [10], one has to sample on average $1/0.2473^8 \approx 70,000$ minimal subsets to hit a *single* promising hypothesis from that structure!

Due to the widespread usage of robust estimators in computer vision many innovations [5], [12], [25], [3], [17], [21] have been proposed to speed-up random hypothesis generation. These methods aim to guide the

• *The authors are with the Sch. of Comp. Sci., The University of Adelaide, Australia, and the Australian Centre for Visual Technologies (ACVT). Phone: +61-8-8313-6188, Fax: +61-8-8313-4366 E-mail: {tjchin, jin.yu, dsuter}@cs.adelaide.edu.au*

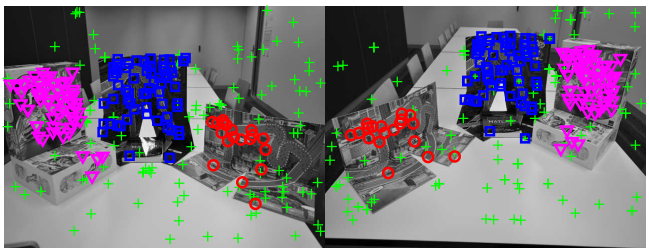


Fig. 1: The "Board Game" image pair with 3 genuine instances of fundamental matrices. Correspondences are established using SIFT [14] matching. False matches are marked with green crosses, while the other markers delineate inlier membership to distinct structures.

sampling process such that the probability of hitting an all-inlier subset is improved. The trick is to endow each input datum with a prior probability of being an inlier and to sample such that data that have high probabilities are more likely to be *simultaneously* selected. Such prior probabilities are often derived from application-specific knowledge. For example Guided-MLESAC [25] and PROSAC [3] concentrate the sampling effort on correspondences with higher keypoint matching scores, the rationale being that inlier correspondences originate from more confident keypoint matches (recall that in geometry estimation a datum consists of a pair of matching points in two views). Proximity sampling assumes that inliers form dense clusters [12] or lie in meaningful image segments [17]. SCRAMSAC [21] imposes a spatial consistency filter so that only correspondences which respect local geometry get sampled.

However, a crucial deficiency of previous methods lies in treating the inlier probability of a datum to be *independent* of the other data. This is untrue when there are multiple structures. Given that a datum $\boldsymbol{x}_i$ is chosen (and thus regarded as an inlier), the probability that $\boldsymbol{x}_j$ is also an inlier (and thus should be chosen as well into the same minimal subset) depends on whether $\boldsymbol{x}_j$ arose from the *same* structure as $\boldsymbol{x}_i$. In other words, it is very possible that two correspondences with high keypoint matching scores are inliers from *different* valid structures. Methods that ignore this point are bound to wastefully generate many invalid *cross-structure* hypotheses. Our results on real and synthetic data (see Fig. 2) prove that this indeed occurs in many previous methods.

Assuming independence on minimal subset selection also precludes the usage of *conditional* sampling strategies in hypothesis generation. More specifically, given that $\tilde{p}$ data with $\tilde{p} < p$ have been chosen into a minimal subset, we should exploit the information available in these $\tilde{p}$ data (and not merely the prior inlier probabilities) to guide the selection of the next datum. The key is to derive and update conditional inlier probabilities (or proxies of these) that allow us to "home into" the more promising data. Such a strategy can greatly improve the chances of retrieving all-inlier minimal subsets from each genuine structure, and is essential for higher order models where one cannot depend on pure random selection or rough prior inlier probabilities. Moreover conditional sampling also encourages data selection within coherent structures. However most previous methods do not sample conditionally, thus they are largely dependent on brute speed to retrieve all-inlier minimal subsets.

In addition we also argue that the domain knowledge used in previous guided sampling techniques may not translate into convincing prior inlier probabilities. For example, false correspondences can have high matching scores, *e.g.* on scenes with repetitive textures many salient patches look similar but these could result in matches that are false according to the epipolar geometry [21], [23]. The idea that inliers form dense clusters also becomes ineffective under high outlier rates as outliers start to encroach the neighbourhood of inliers. In the general case it is questionable whether some usable and reliable domain knowledge is always available.

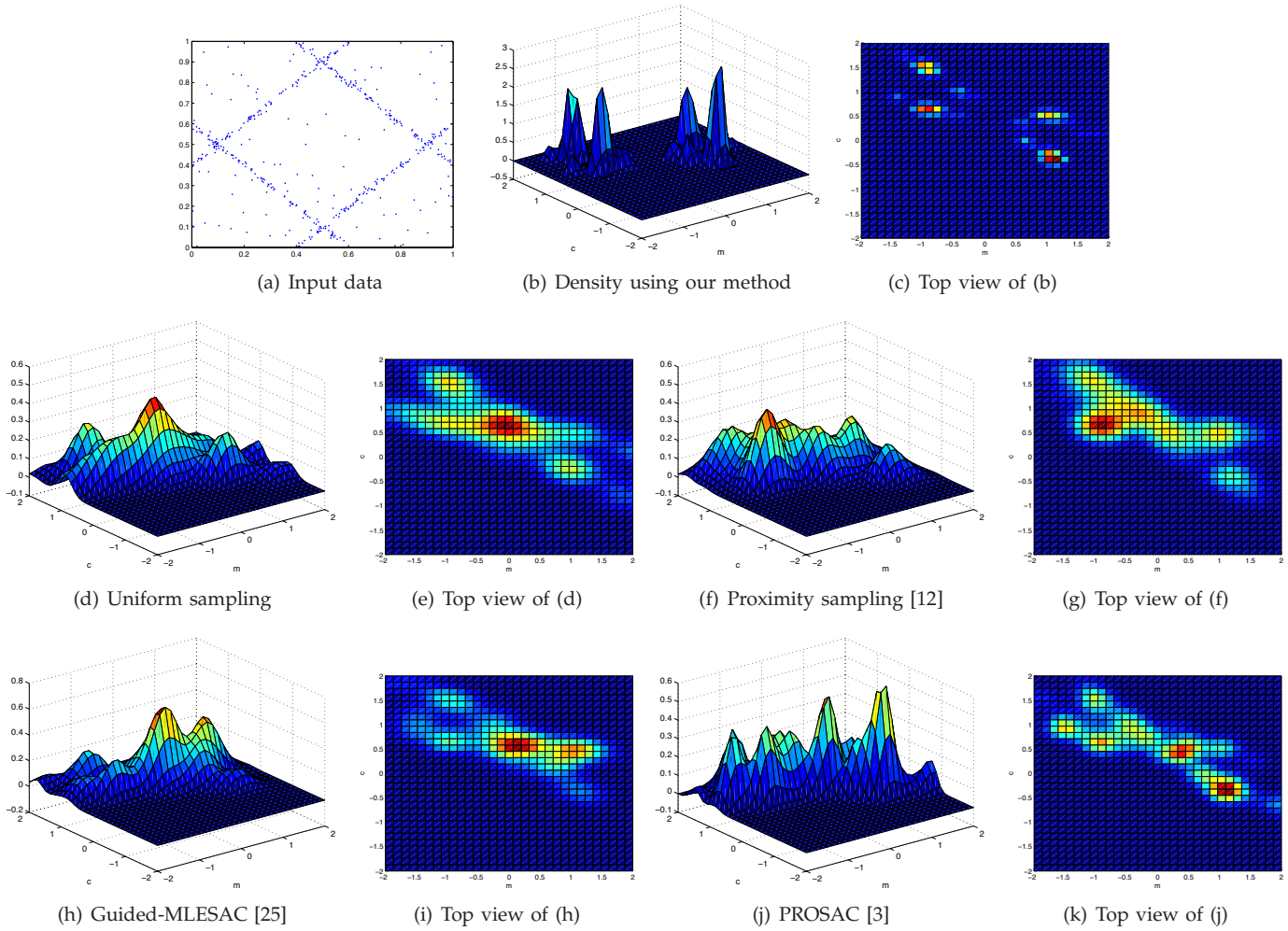
Our contributions. We propose a fundamentally novel technique to accelerate hypothesis generation for robust model fitting. Our guided sampling scheme is driven only by *residual sorting information* and does not require domain- or application-specific knowledge. The scheme is encoded in a series of inlier probabilities estimates which are updated on-the-fly. Most importantly our inlier probabilities are conditioned on the selected data and this enables accurate hypothesis sampling. Under multi-structure data this also encourages sampling within coherent structures. Our technique dramatically reduces the number of samples required, and hence is vastly superior to other sampling methods. In fact we show many cases where other methods simply breakdown while our technique still produces satisfactory results. This work is an extension of our prior work in [2].

The rest of the paper is organised as follows: Sec. 2 surveys related work to put this paper in the right context. Sec. 3 describes the basic principles leading to our novel hypothesis generation scheme in Sec. 4. Sec. 5 provides extensive experimental results and Sec. 6 draws conclusions and discusses future work.

2 Previous Work

Many previous enhancements on the hypothesise-and-verify framework occur in the context of RANSAC. A recent survey [19] categorises them into three groups.

Group 1: Sampling enhancement. The first group of methods aim to improve the random hypothesis sampling routine such that the chances of hitting an all-inlier sample is increased. In [5] the LO-RANSAC method introduces an inner RANSAC loop into the main RANSAC body such that hypotheses may be generated from the set of inliers found so far, thus improving the consensus score more rapidly. A crucial ingredient for the method, however, is the inlier threshold for distinguishing inliers from outliers. In cases with complex geometric models, high-dimensional data and intricate



(a) Input data (b) Density using our method (c) Top view of (b)

(d) Uniform sampling (e) Top view of (d) (f) Proximity sampling [12] (g) Top view of (f)

(h) Guided-MLESAC [25] (i) Top view of (h) (j) PROSAC [3] (k) Top view of (j)

Fig. 2: Given the input data in (a) where there are 4 lines with 100 points per line and 100 gross outliers, we sample 100 line hypotheses (each estimated from a minimal subset of size 2) using the proposed method, uniform sampling (à la the original RANSAC [8]), proximity sampling [12], Guided-MLESAC [25] and PROSAC [3], yielding the parameter space density plotted respectively in (b), (d), (f), (h) and (j). Notice that the hypotheses of our method are concentrated on the correct models yielding 4 distinct peaks. Results from the other methods, however, contain many false peaks and irrelevant hypotheses. As Sec. 5 shows, in multi-structure data our method can successfully "hit" all true models using considerably less time than other methods.

residual functions, considerable tuning effort is required to obtain an appropriate inlier threshold.

Guided-MLESAC [25] and PROSAC [3] focus the sampling on more confident keypoint matches. These methods, however, are essentially guided only by the prior inlier probabilities and do not conduct conditional sampling to further improve efficiency. The reliance on prior inliers probabilities also renders them prone to sampling invalid cross-structure hypotheses given multi-structure data. Moreover, scenes with repetitive textures may give rise to correspondences that are well-matched (*i.e.*, high scores) in terms of local appearance, but incorrect according to the epipolar geometry [21], [23].

In [12] sampling is concentrated on neighbouring correspondences, and in a similar spirit GroupSAC [17] focusses sampling on groups of data obtained using image segmentation. SCRAMSAC [21] conducts a spatial filtering step such that matches with similar local geometry are considered. These methods thus make the assumption that inliers from the same structure cluster together in the image (spatial) domain. We show later that this assumption is violated and the methods are easily confused in richly textured scenes where outliers co-exist densely with inliers in local neighbourhoods. Moreover the notion of proximity usually requires a scale or bandwidth parameter (e.g. σ for Gaussian neighbourhood) that can be difficult to tune for some input spaces.

We emphasise that our work belongs to this group. However in contrast to the other methods in this group we construct and update on-the-fly conditional inlier probabilities for minimal subset selection. This supports highly efficient hypothesis sampling especially for higher-order models and multi-structure data. Further our method is domain-independent and does not require

potentially confusing prior inlier probabilities or calculations of data proximity measures.

Group 2: Efficient verification. This group of methods speed-up hypothesis verification by minimising the time expended for evaluating unpromising hypotheses. The $T_{d,d}$ test [15] evaluates a hypothesis on a small random subset of the input data. This may mistakenly reject good hypotheses thus a much larger number of samples are required. However the overall time can potentially be reduced since the verification now consumes less time. The Bail-Out test [1] and WaldSAC [16], [4] respectively apply catch-and-release statistics and Wald's theory of sequential decision making to allow early termination of the verification of an unfavourable hypothesis.

Group 3: Optimising order of verification. The third group aims for real-time RANSAC. The task is to find the best model from a fixed number of hypotheses afforded by the time interval. Given a set of hypotheses, Preemptive RANSAC [18] scores them in a breadth-first manner such that unpromising hypotheses can be quickly filtered out from the subsequent passes. ARRSAC [19] performs a partially breadth-first verification such that the number of hypotheses may be modified according to the inlier ratio estimate while still bounding the runtime.

We are also aware of recent work [7], [13] that sidesteps the hypothesise-and-verify framework and solves robust estimation directly as a global optimisation problem. While providing globally optimal solutions, these methods require considerably more time than RANSAC. Our aim here is to efficiently fit a geometric model onto noisy data with minimal loss to accuracy, and therefore our aims are different to [7], [13]. We also note that these methods [7], [13] currently cannot handle multi-structure data which are prevalent in practical applications.

3 Inlier Probabilities from Sorting

We first describe how inlier probabilities can be estimated from residual sorting information. Let $\mathcal{X} := \{\boldsymbol{x}_i\}_{i=1}^{N}$ be a set of N input data. Under the hypothesise-and-verify framework, a series of tentative models (or hypotheses) $\{\theta_1, \ldots, \theta_M\}$ are fitted on randomly selected minimal subsets of the input data where M is the number of hypotheses generated. For each datum $\boldsymbol{x}_i$ we compute its absolute residuals as measured to the M hypotheses to form the residual vector

$$\boldsymbol{r}^{(i)} := [\ r_1^{(i)}\ r_2^{(i)}\ \cdots\ r_M^{(i)}\ ]. \tag{1}$$

Note that the hypotheses do not lie in any particular order except the order in which they are generated. We then find the permutation

$$\boldsymbol{a}^{(i)} := [\ a_1^{(i)}\ a_2^{(i)}\ \cdots\ a_M^{(i)}\ ] \tag{2}$$

such that the elements in $\boldsymbol{r}^{(i)}$ are sorted in non-descending order, *i.e.*,

$$u < v \Longrightarrow r^{(i)}_{a_u^{(i)}} \leq r^{(i)}_{a_v^{(i)}}. \tag{3}$$

The sorting $\boldsymbol{a}^{(i)}$ essentially ranks the M hypotheses according to the *preference* of $\boldsymbol{x}_i$; the higher a hypothesis is ranked the more likely $\boldsymbol{x}_i$ is an inlier to it.

Intuitively, two data $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ will share many common hypotheses at the top of their preference list $\boldsymbol{a}^{(i)}$ and $\boldsymbol{a}^{(j)}$ *if* they are inliers from the *same* structure. This is independent of whether $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are, for example, correspondences with high keypoint matching scores. To elicit this phenomenon, first let $\boldsymbol{a}^{(i)}_{1:h}$ be the vector with the first-h elements of $\boldsymbol{a}^{(i)}$. We define the following function as the "intersection" between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$:

$$f(\boldsymbol{x}_i, \boldsymbol{x}_j) := \frac{1}{h}\left|\boldsymbol{a}^{(i)}_{1:h} \cap \boldsymbol{a}^{(j)}_{1:h}\right|, \tag{4}$$

where $|\boldsymbol{a}^{(i)}_{1:h} \cap \boldsymbol{a}^{(j)}_{1:h}|$ finds the number of common elements shared by $\boldsymbol{a}^{(i)}_{1:h}$ and $\boldsymbol{a}^{(j)}_{1:h}$. Window size h with $1 \leq h \leq M$ specifies the number of leading hypotheses to take into account. Note that $f(\boldsymbol{x}_i, \boldsymbol{x}_j)$ ranges between 0 and 1 and is symmetric. Also $f(\boldsymbol{x}_i, \boldsymbol{x}_i) = 1$ for all i.

We demonstrate this effect using the data in Fig. 1. Global coordinate normalisation is first performed on the correspondences following [10]. We then generate $M = 1000$ fundamental matrix hypotheses, each fitted using Direct Linear Transformation (DLT) [11] on a minimal subset of size $p = 8$ chosen via *pure* random selection. The residual $r_m^{(i)}$ is taken as the Sampson distance of datum $\boldsymbol{x}_i$ to the m-th hypothesis. We then obtain the responses of $f(\boldsymbol{x}_i, \boldsymbol{x}_j)$ for *all* unique pairs of the input data while h is varied from $1, \ldots, M$. Fig. 3(a) plots the average of the responses which are separated into two sets: Set "SS" contains pairs of inliers from the same structure while set "DS" contains the rest. The result clearly shows that inliers from the same structure have higher intersection values relative to other possible pairs of inputs. Also, the gap between SS and DS rapidly grows as h is increased (under pure random hypotheses sampling, the expected response from DS can be shown to be linear against h). It quickly attains the maximum value before decreasing slowly to 0 at $h = M$. This indicates that h controls the *discriminative power* of $f(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

We then obtain the $N \times N$ matrix K where the element at the i-th row and j-th column is $f(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Fig. 3(b) displays K for the data in Fig. 1 when $h = 100$ (or $h = 0.1M$) for $f(\boldsymbol{x}_i, \boldsymbol{x}_j)$. The points are arranged according to their structure membership, *i.e.*, $\boldsymbol{x}_1$ to $\boldsymbol{x}_{69}$ are inliers from structure 1, $\boldsymbol{x}_{70}$ to $\boldsymbol{x}_{137}$ are inliers from structure 2 and so on. The gross outliers are $\boldsymbol{x}_{167}$ to $\boldsymbol{x}_{279}$. This makes visible a block diagonal pattern which confirms that strong mutual support occurs among inliers of the same structure. We emphasise that such a data arrangement is purely to aid in presentation and is unnecessary for $f(\boldsymbol{x}_i, \boldsymbol{x}_j)$, or for our subsequent steps to work.

We further analyse the results by averaging the rows of K according to the above data arrangement. Fig. 3(c) shows the result. Unsurprisingly, the results show that on average the row values for an inlier concentrate mostly on other inliers from the *same* structure, while

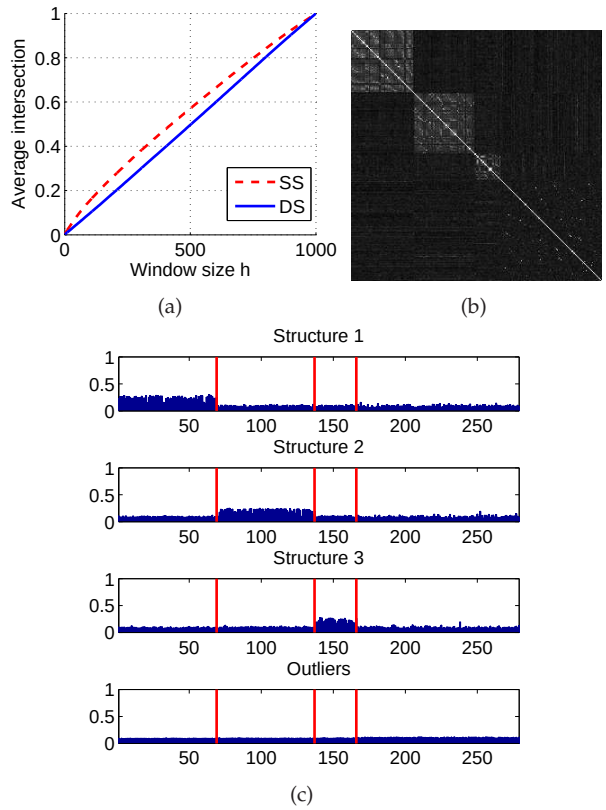


Fig. 3: (a) Average intersection values for the Board Game image pair (Fig. 1) for varying h. SS: Same structure, DS: Otherwise. (b) The corresponding matrix K of size 279×279 with parameter h in $f(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is set to 100 (or $h = 0.1M$). (c) Average values of rows of matrix K grouped according to structure membership. Red vertical lines delineate structure/group boundary.

for a gross outlier the row values are generally low and appear to be randomly distributed.

Our idea is to use these intersection values to drive conditional sampling of minimal subsets. Given that a datum is selected, the intersection values of that datum with the rest of the data yield sampling weights (*i.e.*, a useful proxy for conditional inlier probabilities) to seek the second datum. The intersection values of the first and second data then provide guidance to pick the third datum, and so on (Sec. 4 explains how weights are combined). This results in a sampling scheme that encourages minimal subset selection within consistent structures. Our derivations are general and equally applicable to a wide range of computer vision problems.

Interestingly, this phenomenon of coherence in preference among inliers from the same structure is realised without obtaining a single all-inlier minimal subset from *any* of the true model instances in the data. Indeed, for the Board Game image pair, it is statistically unlikely that even one all-inlier minimal subset exists among the 1000 randomly chosen samples used in Fig. 3 (recall that we would need on average $\approx 70,000$ samples before retrieving one all-inlier minimal subset for structure 1).

4 Accelerated Hypothesis Generation

In this section we describe how residual sorting can be exploited to drive a very efficient hypothesis generation scheme for robust model fitting.

4.1 Calculating conditional inlier probabilities

Assume M model hypotheses $\{\theta_1, \ldots, \theta_M\}$ have been generated so far and we wish to sample the next hypotheses (in a guided fashion). Let the model to be fitted be determined by a minimal subset

$$\mathcal{S} = \{\boldsymbol{x}_{s_1} \ldots \boldsymbol{x}_{s_p}\} \subset \mathcal{X} \tag{5}$$

of size p, where

$$\{s_1, \ldots, s_p\} \subset \{1, \ldots, N\} \tag{6}$$

is a set of indices identifying the members of $\mathcal{X}$ present in $\mathcal{S}$. Sampling a hypothesis is thus equivalent to determining the values of $\{s_j\}_{j=1}^{p}$. Henceforth, where no confusion arises $\boldsymbol{x}_{s_j}$ and s_j are used interchangeably.

The first datum s_1 is selected purely randomly, *i.e.*, s_1 is sampled from the discrete uniform distribution

$$s_1 \sim U(1, N). \tag{7}$$

To select the second datum s_2 we construct the conditional inlier probability distribution

$$P_1(i) := \begin{cases} \alpha_1 f(\boldsymbol{x}_i, \boldsymbol{x}_{s_1}) & \text{if } \; i \neq s_1, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where f is the intersection function (4) based on the M hypotheses sampled thus far. Normalisation constant α_1

$$\alpha_1 = \frac{1}{\sum_{i \neq s_1} f(\boldsymbol{x}_i, \boldsymbol{x}_{s_1})} \tag{9}$$

ensures that $P_1(i)$ is a valid discrete probability distribution, while imposing $P_1(i) = 0$ for $i = s_1$ guarantees sampling without replacement. The second datum s_2 is then selected according to P_1, *i.e.*,

$$s_2 \sim P_1. \tag{10}$$

This can be accomplished by using the values of P_1 as a set of *sampling weights*, *i.e.*, if

$$P_1(u) > P_1(v) \tag{11}$$

then $\boldsymbol{x}_u$ is more likely to be selected than $\boldsymbol{x}_v$.

We now wish to combine the information provided by s_1 and s_2 to sample s_3, and in general to use the information provided by the k data chosen thus far to sample the $(k+1)$-th datum. To this end we construct the k-th conditional inlier probability distribution as

$$P_k(i) := \begin{cases} \alpha_k \prod_{j=1}^{k} f(\boldsymbol{x}_i, \boldsymbol{x}_{s_j}) & \text{if } \; i \notin \{s_1, \ldots, s_k\}, \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

and sample

$$s_{k+1} \sim P_k. \tag{13}$$

Again, analogously to Eq. (9), α_k normalises P_k to maintain it as a valid discrete probability distribution.

Intuitively, (12) is equivalent to the element-wise multiplication of rows $s_1, \ldots, s_k$ of matrix K with the s_1-th, $\ldots$, s_k-th elements set to zero. The guidance provided by P_k, however, is contingent on window size h in $f(\boldsymbol{x}_i, \boldsymbol{x}_j)$. As discussed in Sec. 3, h controls the discriminative power of $f(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Fig. 3(a) as well as further experimentations in Sec. 5.2 lead us to conclude that a large range of h, i.e. $\lceil 0.05M \rceil \leq h \leq \lceil 0.4M \rceil$, is effective for the task. In all our results in Sec. 5 h is set to $\lceil 0.1M \rceil$. Note also that if $h = M$, $f(\boldsymbol{x}_i, \boldsymbol{x}_j) = 1$ for all i, j thus the P_k's reduce to the discrete uniform distribution.

4.2 Bootstrap effect on hypothesis sampling

We show how the sampling weights in Eq. (12) achieve a "bootstrap" effect in hypothesis sampling. A small initial set of hypotheses, sampled practically randomly, provide rough guidance to select the next minimal subsets. The newly fitted hypotheses then improve the accuracy of the conditional inlier probabilities for retrieving other useful minimal subsets. The result is that the desired block diagonal pattern in K occurs at a much lower M.

Fig. 4 depicts an actual run of weighted sampling on the Board Game image pair. By the 25-th iteration, sampling weights good enough for differentiating inliers from most of the outliers, as well as a rough block diagonal pattern, have emerged. By only the 50-th iteration sampling weights for dichotimising different structures have materialised. The result at this stage is already visibly better than the result of random sampling at 1000 iterations (c.f. Fig. 3). Fig. 5 shows the outcome after 400 iterations of conditional sampling, where the existence of three structures in K is even more obvious. Moreover, by the 400-th iteration at least one all-inlier minimal subset have been retrieved for each structure; this represents a massive improvement in efficiency. Using pure random selection and referring to the smallest structure in the data (10.39% inliers), on average $1/0.1039^8 \approx 73,000,000$ iterations are required before hitting a single all-inlier minimal subset from each structure!

As alluded to before this bootstrap effect is achieved *without* having sampled a single all-inlier minimal subset in the initial stages (the first-four histograms in the 3rd row of Fig. 4 are thus empty). This departs from the intuition that at least one true hypothesis from each structure is required. Our result is clear empirical evidence that seemingly "worthless" (according to RANSAC criterion) hypotheses do provide useful information for structure dichotomy, since all that is required is for inliers from the same structure mutually prefer the same hypotheses.

4.3 Efficient implementation

After a new hypothesis is sampled the conditional inlier probabilities (12) can be updated by first appending the

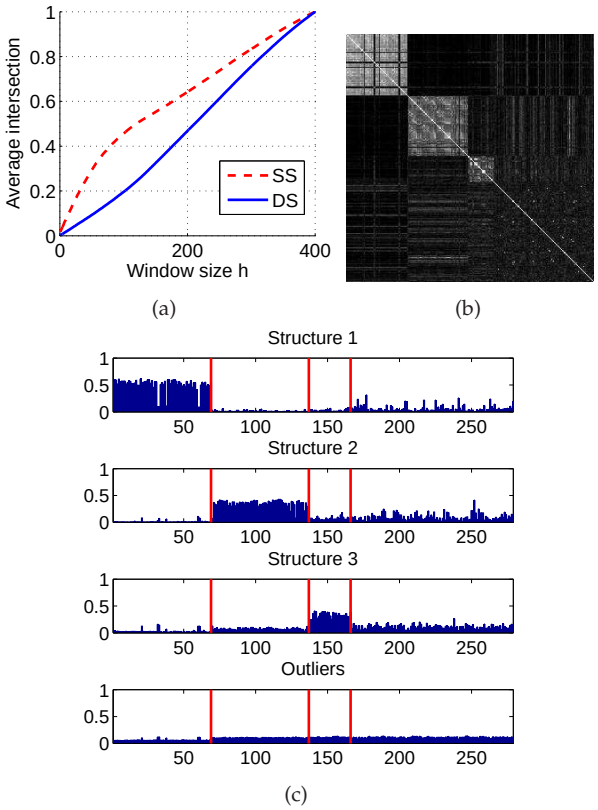


Fig. 5: Results after 400 iterations of sampling from the conditional inlier probabilities (12) for the Board Game image pair. (a) Avg. intersection values. (b) K matrix. (c) Avg. of rows of K according to structure membership.

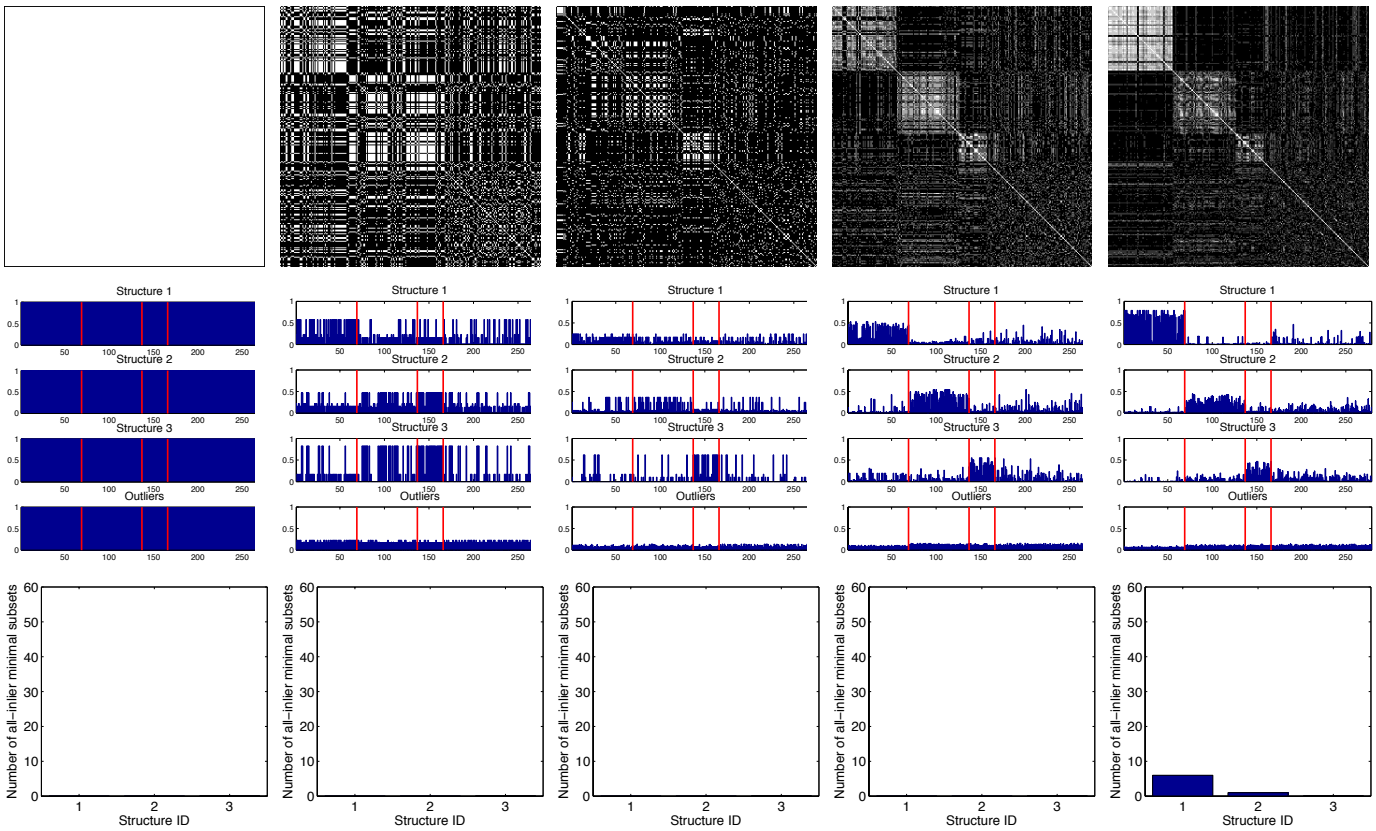
absolute residuals as measured to the new hypothesis to the residual vector of each datum (1), i.e.

$$\boldsymbol{r}_{new}^{(i)} = [\ \boldsymbol{r}_{old}^{(i)}\ \ r_{M+1}^{(i)}\ ], \quad \text{where } \boldsymbol{r}_{old}^{(i)} \in \mathbb{R}^{1\times M}, \qquad (14)$$

and then re-sorting $\boldsymbol{r}_{new}^{(i)}$ to obtain the new permutation $\boldsymbol{a}_{new}^{(i)}$. Window size h is updated accordingly as $h_{new} = \lceil 0.1(M+1) \rceil$. The efficiency of the update can be improved by maintaining the already sorted $\boldsymbol{r}_{old}^{(i)}$ in a data structure that facilitates efficient search and insertion for $r_{M+1}^{(i)}$, e.g. binary trees with $\mathcal{O}(\log M)$ complexity for the operations. More sophisticated implementations, however, are only worthwhile for very large M's (in the order of 10^6). In most of our cases $M < 10^4$ and $\boldsymbol{r}_{new}^{(i)}$ can simply be wholly re-sorted with quicksort without noticeable increases in computational overheads.

Furthermore, updating the residual sorting as soon as a new hypothesis arrives is unnecessarily conservative because a single new hypothesis does not add much information about inlier probabilities. Our proposed algorithm thus updates the sorting permutation $\boldsymbol{a}^{(i)}$ and advances h only after a block (of size b) of new hypotheses are generated. Algorithm 1 provides the pseudocode for our method which we call "Multi-GS". Sec. 5.2 explores how h and b affect sampling efficiency.

More careful thought, however, should be put into



(a) Iteration 1 (b) Iteration 5 (c) Iteration 10 (d) Iteration 25 (e) Iteration 50

Fig. 4: Boostrapping hypothesis sampling for the Board Game image pair. Results after 1, 5, 10, 25 and 50 iterations (respectively columns 1–5) of sampling using the proposed conditional inlier probabilities (12) are shown. Top row: Resulting matrix K of intersection values. Middle row: Average row values of K separated according to structure membership. Bottom row: Count of all-inlier minimal subsets retrieved for each structure. Zero all-inlier minimal subsets are sampled in the initial stages, hence *the first-four histograms are empty*.

implementing the intersection required in $f(\boldsymbol{x}_i, \boldsymbol{x}_j)$ (4). A straightforward algorithm with a nested loop will scale as $\mathcal{O}(h^2)$. Using a symbol table [22] reduces this to a more efficient $\mathcal{O}(h)$ as described in Algorithm 2.

On the surface it seems that the somewhat involved computation is a weakness. However our algorithm conducts a more informed sampling given a unit of computation time in comparison to other techniques. The result is that we require less effective CPU time to hit all-inlier minimal subsets, especially for models of higher order and heavily contaminated data.

5 Experiments

We evaluate the performance of the proposed method (Multi-GS, Algorithm 1) on common computer vision tasks. We compared against the following state-of-the-art sampling methods: (1) LO-RANSAC [5], (2) Proximity sampling [12], (3) Guided-MLESAC [25], and (4) PROSAC [3]. We consider pure random sampling as in the original RANSAC [8] as the baseline. We implemented all algorithms in MATLAB. All experiments were run on a Linux machine with a 3.0GHz Intel Core 2 Duo processor and 4 GB of main memory.

Parameter settings for all methods are detailed in Table 1. Note that the inlier threshold for LO-RANSAC is obtained using ground truth knowledge of the inlier-outlier identity of each datum; This is an unfair advantage since the method is allowed to access the ground truth. We also stress that, unless stated otherwise, for Multi-GS we consistently set $h = \lceil 0.1M \rceil$ and $b = 10$. b.

Method/parameter	Setting
LO-RANSAC	
- inlier threshold	median of inlier residuals
Proximity sampling	
- std. dev. of normal dist.	2 times mean pairwise distance
Guided-MLESAC	
- prior inlier probabilities	matching scores (if available)
PROSAC	
- T_N	2.5×10^4
- prior inlier probabilities	matching scores (if available)
Multi-GS (proposed)	
- Window size h	$\lceil 0.1M \rceil$
- Block size b	10

TABLE 1: Parameter settings for all methods.

Algorithm 1 Accelerated Multi-Structure Hypothesis Generation by Preference Analysis (Multi-GS)

input input data $\mathcal{X}$, total required number of hypotheses T, size of a minimal subset $p > 0$ and block size $b > 0$
output a set Θ of T model hypotheses
for $M := 1, 2, \ldots, T$ **do**
  if $M \leq b$ **then**
    randomly sample p data and store as $\mathcal{S}$
  else
    select at random s_1 and initialise $\mathcal{S} := \{\boldsymbol{x}_{s_1}\}$
    for $k := 1, 2, \cdots, (p-1)$ **do**
      construct P_k (12) using $\mathbf{a}^{(i)}$, h and data in $\mathcal{S}$
      sample $\boldsymbol{s}_{k+1}$ according to P_k
      append $\mathcal{S} := \mathcal{S} \cup \{\boldsymbol{x}_{s_{k+1}}\}$
    end for
  end if
  $\Theta = \Theta \cup$ {New hypothesis instantiated from $\mathcal{S}$}
  append to all $\boldsymbol{r}^{(i)}$ abs. residual to new hypothesis
  if $M \geq b$ and $\mathtt{mod}(M, b) = 0$ **then**
    sort all $\boldsymbol{r}^{(i)}$ to obtain permutations $\boldsymbol{a}^{(i)}$
    $h := \lceil 0.1M \rceil$
  end if
end for
return Θ

Algorithm 2 Computing $f(\boldsymbol{x}_i, \boldsymbol{x}_j)$ using symbol tables

input residual sorting indices $\boldsymbol{a}^{(i)} = [a_1^{(i)}\ a_2^{(i)}\ \cdots\ a_M^{(i)}]$ and $\boldsymbol{a}^{(j)} = [\ a_1^{(j)}\ a_2^{(j)}\ \cdots\ a_M^{(j)}\ ]$, desired win. size h
output response of $f(\boldsymbol{x}_i, \boldsymbol{x}_j)$
initialise to zero symbol table $\mathbf{y}$ of size M
for $s := 1, 2, \ldots, h$ **do**
  $\mathbf{y}(a_s^{(i)}) := 1$ (set the $a_s^{(i)}$-th element of $\mathbf{y}$ to 1)
end for
$acc := 0$
for $s := 1, 2, \ldots, h$ **do**
  if $\mathbf{y}(a_s^{(j)}) = 1$ **then**
    $acc := acc + 1$
  end if
end for
return $acc \div h$

5.1 Performance under different outlier rates

We first examine the performance of Multi-GS under different outlier rates. We focus on the fundamental matrix estimation problem for single- and multi-structure data.

Single-structure data. We use the Barr-Smith Library image pair shown in Fig. 8(c). The images are resized to 240×320 pixels and SIFT keypoints are detected and matched across the two views. True and false matches are manually identified. This yields a dataset with 70 inliers (20.71%) and 268 outliers (79.29%). Global coordinate normalisation [10] is performed on the data for numerical stability. The 8-point method [11] is applied to fit fundamental matrices on minimal subsets of size 8. The residual is taken as the symmetric transfer error.

Each run of this experiment involves a data instance which contains all 70 inliers and L outliers, where $L = 26, 52, \ldots, 260$. For each L, 50 data instances are produced, where for each instance the L outliers are randomly selected from the full set of 268 outliers. On each data instance we run all compared methods for a maximum of 10 seconds. We then record

1) the total number of minimal subsets sampled,
2) among the sampled minimal subsets, the number of *all-inlier* minimal subsets produced,
3) the number of steps (i.e. number of minimal subsets sampled) before hitting the *first* all-inlier minimal subset (i.e. the solution), and
4) the elapsed time when the first all-inlier minimal subset is obtained.

For each value of L the median results over the 50 repetitions are obtained and plotted in Fig. 6 (top row).

Fig. 6(a) shows that within the 10s limit Multi-GS is capable of sampling less than 1/3 of the hypotheses generated by the other methods. This stems from the higher computational demands of updating the conditional weights. However, despite the smaller number of sampling steps per unit time, Multi-GS can produce many more all-inlier minimal subsets than the other methods, especially for outlier rates of 20% and above; see Fig. 6(b). In fact, at the highest outlier rate ($\approx 80\%$), only Multi-GS and PROSAC can still produce all-inlier minimal subsets within 10s (with Multi-GS returning a slightly higher number than PROSAC; see Table 2), while the other methods fail to find a single solution.

In terms of the number of iterations and time before hitting the first solution, PROSAC which benefits from accurate keypoint matching scores has an overwhelming advantage. The first all-inlier minimal subset is almost always retrieved by PROSAC in the very first iteration; see Figs. 6(c) and 6(d). Multi-GS is relatively close behind. At the highest outlier rate Multi-GS retrieves the first all-inlier minimal subset in $\approx$ 300 iterations or 0.89s (see Table 2). Only PROSAC and Multi-GS is capable of producing sub-1s performances on the Barr-Smith Library image pair. Note that this experiment is conducted using *single-structure* data. We shall compare the methods on *multi-structure* data next.

Multi-structure data. We repeat the previous experiment, this time using the Board Game image pair which consist of 3 fundamental matrix instances; see Fig. 1. The data contains 113 outliers (40.50%) and 166 inliers (59.5%) with the inliers separated into 3 groups of size 69 (24.73%), 68 (24.37%) and 29 (10.39%). As before multiple data instances are generated, each containing all the inliers and L randomly selected outliers with $L = 11, 22, \ldots, 110$. For each L value 50 repetitions are generated. Considering that the data is multi-structure, we regard the arrival of a solution as the *retrieval of at least one all-inlier minimal subset from each structure*. Fig. 6 (bottom row) illustrates the median results over 50 repetitions. Note that the outlier rates in Fig. 6 (bottom

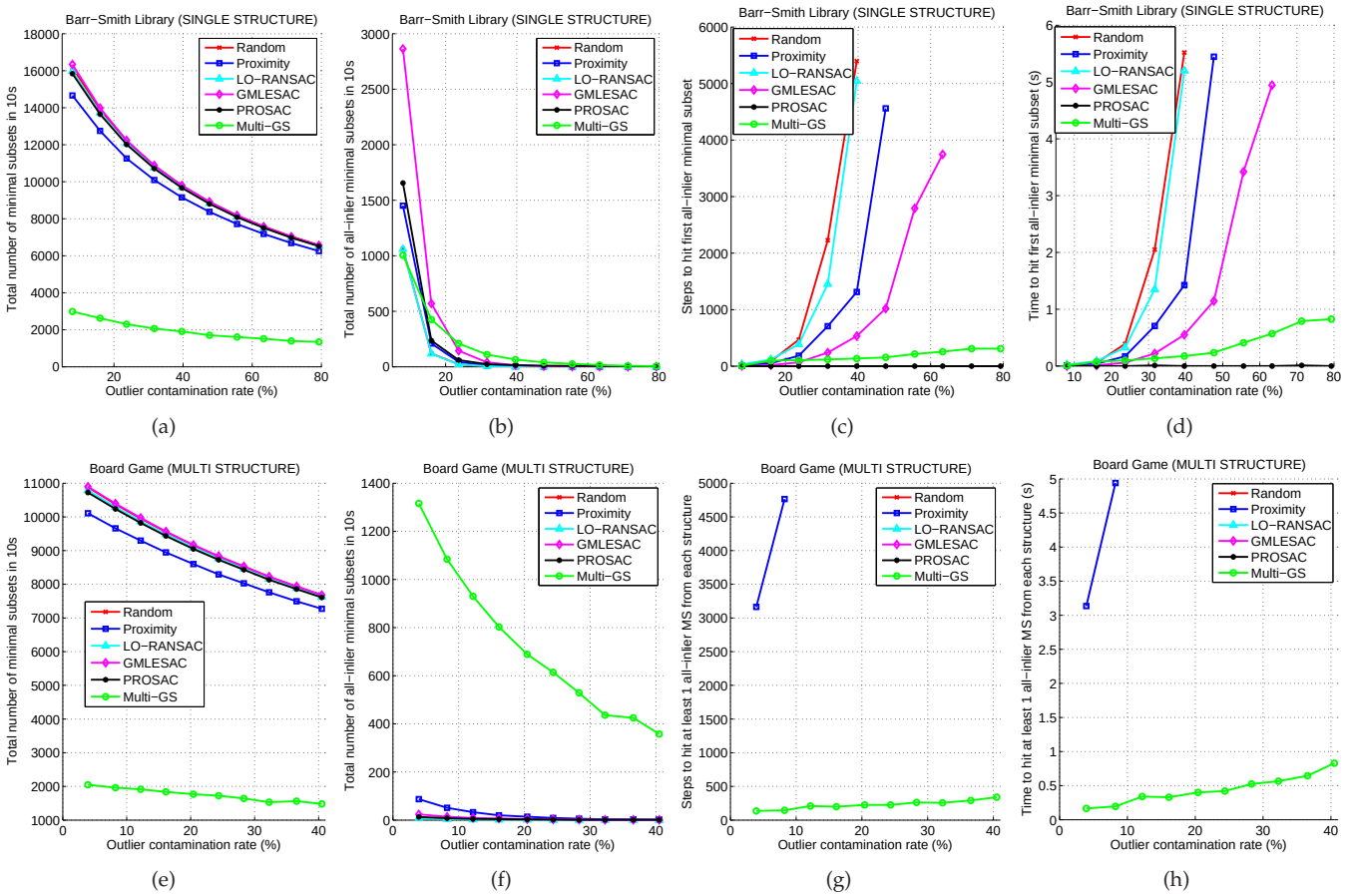


Fig. 6: Performance under different outlier rates (best viewed in colour). Top row: Barr-Smith Library (**single structure**). Bottom row: Board Game (**multi structure**). (a)(e): Number of minimal subsets sampled in 10s. (b)(f): Number of *all-inlier* minimal subsets sampled in 10s. (c)(g): Steps to hit the first solution (not plotted if unable to find solutions within 10s). (d)(h): Time to hit the first solution (not plotted if unable to find solutions within 10s).

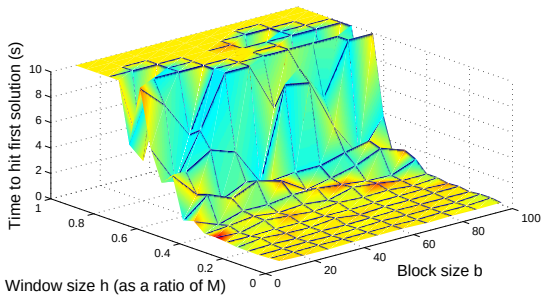
row) is simply taken as $100\% \times L/(L+166)$; the actual per structure outlier rate is actually much higher.

The trends in Fig. 6(e) are similar to those in Fig. 6(a), i.e. Multi-GS is capable of sampling much fewer minimal subsets than the other methods. However, despite performing fewer iterations, as shown in Fig. 6(f) Multi-GS yields overwhelmingly more all-inlier minimal subsets. Since the inliers are clustered Proximity Sampling is capable of yielding a decent amount of all-inlier minimal subsets, but this performance deteriorates rapidly as the outlier rate increases. As expected PROSAC and Guided-MLESAC are unable to retrieve all-inlier minimal subsets due to their tendency to include inliers from *different* structures into the same minimal subset. Keypoint matching scores alone cannot distinguish inliers from different structures. Finally, as can be seen in Figs. 6(g) and 6(h) only Multi-GS can obtain solutions within the 10s limit. In fact, Multi-GS is able of retrieving all-inlier minimal subsets from each structure within 1s.

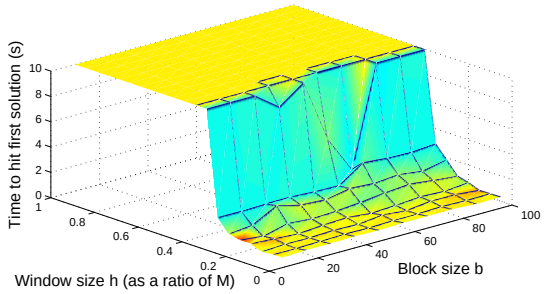
5.2 Effects of parameters h and b

We examine empirically the effects of varying parameters h (window size) and b (block size); see Algorithm 1. We again use the Barr-Smith Library (Fig. 8(c)) and Board Game (Fig. 1) image pairs with 50% of the outliers included. In this experiment we parameterise window size h as a ratio of the number of hypotheses M sampled so far. The two parameters are set in the range of $h = [\ 0.05\ 0.1\ \ldots\ 1.0\ ]$ and $b = [\ 10\ 20\ \ldots\ 100\ ]$. For each combination of h and b we run Multi-GS 10 times. In each run we record the time elapsed to hit the first solution (again, for the multi-structure case this is regarded as the retrieval of at least one all-inlier minimal subset from each structure). Fig. 7 shows the median results.

It is clear that the performance of Multi-GS is tolerant to block size b. Within the range of b ($10 \leq b \leq 100$) similar results are obtained. The effect of h, however, is more prominent. As expected the performance of Multi-GS decays with the increase of h since the intersection function (4) loses discriminative power. The deterioration however is faster on Board Game (multi-structure) since the per-structure outlier rate is much higher, especially on the smallest structure whose inliers make up 10.39% of the data. Thus, the optimal h should be proportional to the inlier rate of a genuine structure. Nonetheless, even on Board Game, Multi-GS demonstrates a large degree



(a) Time to first solution on Barr-Smith Library.



(b) Time to first solution on Board Game.

Fig. 7: Effects of varying the settings of h and b on Multi-GS (see Algorithm 1). Each point on the surface above indicates the time required to hit the first solution given combination of h and b. Note that (a) is on *single-structure* data while (b) is on *multi-structure* data.

of tolerance to h as it is capable of retrieving solutions within 10s given $h \in [0.05M\ 0.4M]$.

5.3 Further results on fundamental matrix fitting

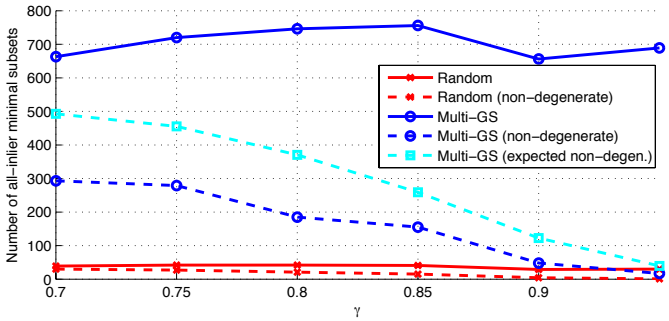
Other image pairs used for fundamental matrix estimation are shown in Fig. 8. Their respective inlier and outlier proportions are listed in Table 2. The experimental setup and performance metric is the same as in Sec. 5.1, except that we do not vary the outlier rates. Table 2 summarises the results obtained. A similar conclusion as that achieved in Sec. 5.1 is apparent here; On single-structure data, PROSAC rapidly retrieves the first solution due to the availability of accurate keypoint matching scores, while Multi-GS is at least on par with the other methods. On multi-structure data, however, all methods except Multi-GS break down. Multi-GS is also capable of consistently retrieving the first solution under 1s, as well as the largest number of all-inlier minimal subsets.

5.3.1 *Performance under degeneracies*

Under the 8-point method, a degenerate fundamental matrix is obtained when more than 6 correspondences among 8 in a minimal subset lie on the same plane [9]. This occurs frequently when there exists a dominant plane in the scene [6]. Degenerate models may have high consensus despite being far from the optimal model.

The Dino-Books data in Fig. 8(f) is modified by removing 2 of the 3 genuine structures; see Fig. 9(a). The inliers

(a) Dino-Books with single fundamental matrix instance (best viewed in colour). The inliers lie on two distinct planes.



(b) Total number of all-inlier minimal subsets produced and, among these, total number of non-degenerate all-inlier minimal subsets produced as the ratio γ is increased (best viewed in colour).

Fig. 9: Performance under degeneracies.

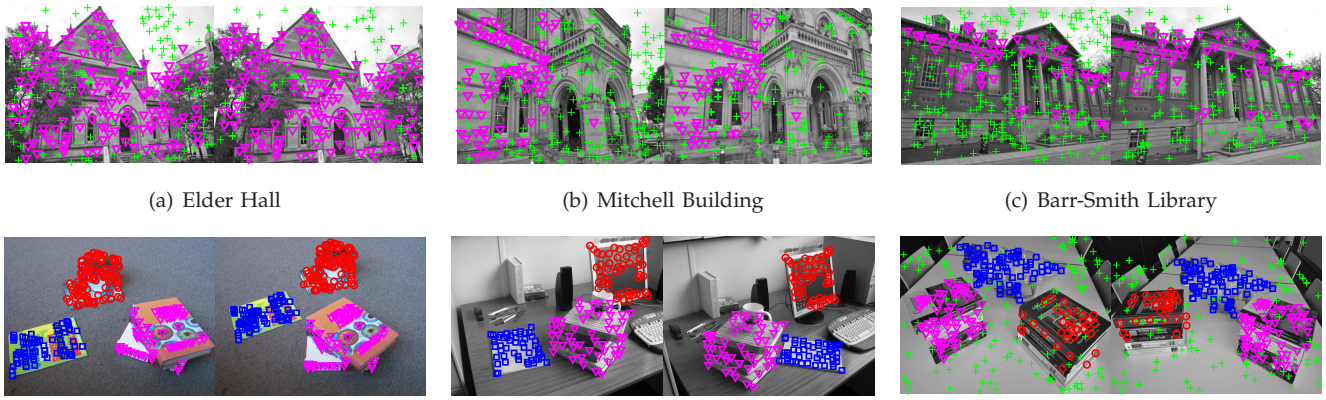
of the remaining structure lie on 2 distinct planes. We indicate these as Set A (magenta triangles, 29 matches) and Set B (blue squares, 49 matches). We create different instances of the data by using all of Set B (regarded as the "dominant plane") while controlling the number of Set A inliers (the "off-plane" inliers) added. This yields a series of ratios of on-plane inliers to total inliers

$$\gamma = |\text{Set B}|/(|\text{Set A}| + |\text{Set B}|). \quad (15)$$

We are interested in the number of non-degenerate hypotheses produced as the above ratio is increased. To this end we vary the above ratio in the range $[0.75\ 0.90]$. For each ratio 50 data instances are generated, where each instance contains the same quantity of outliers as inliers; this maintains the outlier rate at 50%. On each data instance each method is run for 10s. The median results across the 50 repetitions are presented in Fig. 9(b).

Unsurprisingly, since Multi-GS does not explicitly account for degeneracies, the number of non-degenerate hypotheses sampled decreases approx. linearly with γ. Note that in our case since the number of off-plane inliers added |Set A| can exceed 7, a minimal subset with more than 6 members from Set A is also counted as denegerate. Relative to the expected number[1] of non-denegerate hypotheses among all-inlier minimal subsets, it appears that Multi-GS is below the norm, indicating a tendency to sample from data on the dominant plane.

1. First, compute the probability in Eq. (7) of [6], but by replacing ε_H there with γ, and adding a second analogous probability term for the case $|\text{Set A}| \geq 7$. Then multiply the resulting probability with the number of all-inlier minimal subsets produced by Multi-GS.



(a) Elder Hall (b) Mitchell Building (c) Barr-Smith Library

(d) Office (e) Desk (f) Dino-Books

Fig. 8: Image pairs for fundamental matrix estimation. The keypoints are detected and matched using SIFT. The colours indicate group labels while gross outliers are marked as '+'. (a)(b)(c) are **single-structure** while (d)(e)(f) are **multi-structure**. (d)(e) are from http://www.iu.tu-darmstadt.de/datasets while the others are our own data.

However, owing to a much larger number of all-inlier minimal subsets retrieved, Multi-GS is still able to generate significantly more non-degenerate hypotheses. Contrast this with random sampling which produced a relatively insignificant number of non-degenerate all-inlier minimal subsets. Thus in a practical sense, Multi-GS is still a good choice in the presence of degeneracies.

Moreover, it is worth stressing that in prior work [6], [9], degeneracies are not handled in the hypothesis sampling process, *i.e.*, the sampling does not explicitly search for non-degenerate models. Instead a separate mechanism is used to detect degenerate hypotheses (*e.g.* robust rank detection [9]) before invoking a correction routine (*e.g.* plane-and-parallax algorithm [6]) to recover the fundamental matrix from the degenerate hypotheses. Such routines can be easily plugged into Multi-GS to recover non-degenerate hypotheses from the degenerate all-inlier minimal subsets.

5.4 Single- and multi-structure homography fitting

The image pairs used in this experiment are shown in Fig. 10. Each structure in this problem corresponds to a planar surface which can be modelled by a 2D homography. Global coordinate normalisation [10] is conducted on the data before putative homographies are estimated from minimal subsets of 4 matches (i.e. $p = 4$) using DLT [11]. The residual is computed as the symmetric transfer error. On each image pair each method was run 50 times, each run limited to 10s. Within the given time budget the same performance metrics from Sec. 5.1 are recorded. Table 3 summarises the median results.

As shown in Table 3, on almost all image pairs Multi-GS retrieved the highest number of all-inlier minimal subsets. However in this experiment the gap in performance between Multi-GS and the others is not as obvious. This is a direct consequence of a lower order geometric model ($p = 4$), thus the other methods can compensate for their lower accuracy with speed. Nonetheless, the results still demonstrate that Multi-GS is able to obtain a much higher number of all-inlier minimal subsets with much fewer iterations.

The results also reveal the limits of Multi-GS; The highest per structure outlier rate (pseudo and gross outliers) it can tolerate before failing to retrieve all-inlier minimal subsets from a particular structure is $\approx 95\%$.

5.5 Motion subspace fitting

We also consider the affine camera motion segmentation problem. Point trajectories are first tracked across a video sequence and stored in a data matrix

$$\mathbf{D} = [\ \mathbf{d}_1\ \ \ldots\ \ \mathbf{d}_N\ ] \in \mathbb{R}^{2F \times N}, \tag{16}$$

where F is the number of frames and N is the number of trajectories. Each column $\mathbf{d}_i$ corresponds to a particular trajectory. Under the affine camera model columns belonging to the same rigid motion lie on the same 4D subspace. A popular method for motion segmentation is to robustly fit multiple subspaces on the trajectories using RANSAC [26]. Following their work we first project the data matrix onto the first-5 principal subspaces, yielding $\tilde{\mathbf{D}} \in \mathbb{R}^{5 \times N}$. Minimal subsets containing four randomly chosen columns of $\tilde{\mathbf{D}}$ (i.e. $p = 4$) are then produced, and each subset is used to estimated 4D-subspace hypotheses using SVD. The residual $r_m^{(i)}$ is computed as the orthogonal distance of the i-th column to the m-th subspace hypotheses.

We use sequences from the Hopkins 155 benchmark dataset [26] which consists of 155 sequences of tracked feature points. Each sequence contains either 2 or 3 distinct motions (e.g. moving cars, checkerboards) established from 150–300 individual trajectories. See [26] for details of the collection methodology. Note that in these sequences missing or bad tracks were simply discarded, hence from the point of view of robust estimation there are *no gross outliers*; Only pseudo-outliers exist due to

TABLE 2: Performance of various sampling methods for fundamental matrix estimation on image pairs in Fig. 8. Each experiment was given 50 random runs, each for 10 CPU seconds. We report the total number of samples found within the given time budget and the median of CPU seconds (*resp.* sampling steps) that are required to find *at least one* all-inlier minimal subset for *each* structure. If this is not achieved within the time limit a "$\times$" is marked at the corresponding row and column. The median number of all-inlier samples found is listed separately for each structure (Structure-i, i = $1, 2, \cdots$). The number of inliers and the inlier ratio for each structure is given in the parenthesis. The top result with respect to each performance measure are boldfaced.

Data		Random	Proximity	LO-RANSAC	Guided-MLESAC	PROSAC	Multi-GS
Elder Hall Fig. 8(a)	Number of samples	**6527**	6228	6499	6516	6484	1317
	Structure-1 (190, 55.7%)	58	75	66	**380**	321	257
	Sampling Steps	69	57	127	14	**1**	69
	CPU seconds	0.11	0.09	0.20	0.02	**0.00**	0.11
Mitchell Building Fig. 8(b)	Number of samples	**6448**	6166	6430	6470	6419	1300
	Structure-1 (92, 26.4%)	0	1	0	4	51	**72**
	Sampling Steps	$\times$	5499	$\times$	1385	**1**	124
	CPU seconds	$\times$	9.00	$\times$	2.14	**0.00**	0.24
Barr-Smith Library Fig. 8(c)	Number of samples	**6568**	6256	6544	6561	6522	1343
	Structure-1 (70, 20.7%)	0	0	0	0	5	**7**
	Sampling Steps	$\times$	$\times$	$\times$	$\times$	**1**	319
	CPU seconds	$\times$	$\times$	$\times$	$\times$	**0.00**	0.87
Office Fig. 8(d)	Number of samples	**7124**	6763	7086	7116	7067	1402
	Structure-1 (127, 41.4%)	5	59	5	6	5	**412**
	Structure-2 (85, 27.7%)	0	11	0	0	0	**286**
	Structure-3 (95, 30.9%)	1	13	0	0	0	**281**
	Sampling Steps	$\times$	842	$\times$	$\times$	$\times$	**124**
	CPU seconds	$\times$	1.24	$\times$	$\times$	$\times$	**0.21**
Desk Fig. 8(e)	Number of samples	**11686**	10804	11595	11712	11512	2250
	Structure-1 (50, 31.3%)	0	11	0	0	0	**488**
	Structure-2 (50, 31.3%)	1	22	1	1	5	**611**
	Structure-3 (60, 37.5%)	3	87	4	5	3	**618**
	Sampling Steps	$\times$	940	$\times$	$\times$	$\times$	**117**
	CPU seconds	$\times$	0.87	$\times$	$\times$	$\times$	**0.12**
Board Game Fig. 1	Number of samples	**7693**	7279	7650	7685	7618	1490
	Structure-1 (69, 24.7%)	0	1	0	0	0	**219**
	Structure-2 (68, 24.4%)	0	0	0	0	0	**131**
	Structure-3 (29, 10.4%)	0	0	0	0	0	**11**
	Sampling Steps	$\times$	$\times$	$\times$	$\times$	$\times$	**310**
	CPU seconds	$\times$	$\times$	$\times$	$\times$	$\times$	**0.73**
Dino-Books Fig. 8(f)	Number of samples	**6230**	5944	6203	6218	6183	1348
	Structure-1 (78, 17.7%)	0	0	0	0	0	**98**
	Structure-2 (86, 23.9%)	0	0	0	0	0	**72**
	Structure-3 (41, 11.4%)	0	0	0	0	0	**22**
	Sampling Steps	$\times$	$\times$	$\times$	$\times$	$\times$	**263**
	CPU seconds	$\times$	$\times$	$\times$	$\times$	$\times$	**0.71**

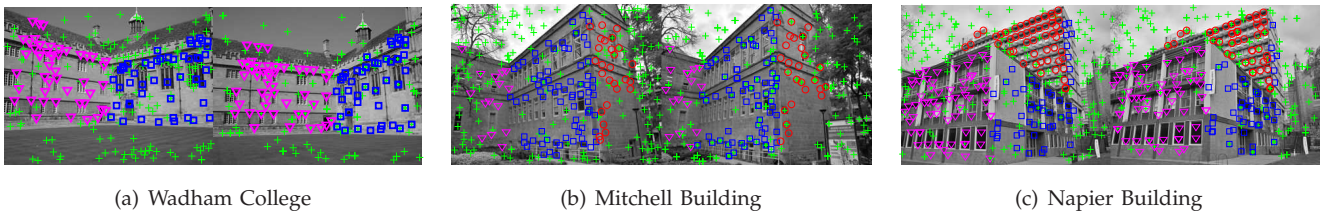
the multi-structure data. Therefore this dataset is significantly easier than the data considered previously.

We compare Multi-GS against the methods listed in Table 1. The parameters (which differ from Table 1) are:

1) For Guided-MLESAC and PROSAC, since there are no gross outliers the prior inlier probabilities (i.e. data quality scores) are set to $1/N$ for all datum.
2) For proximity sampling the std. deviation is taken as twice the average inter-datum distance in $\mathbb{R}^5$.

On each of the 155 sequences each method is given 20 runs, each run limited to 10 seconds. The same performance metrics for the previous multi-structure data experiments are used here. The median of each metric is plotted in Fig. 11. Note that in this dataset we are unable to smoothly vary the outlier rate like in Fig. 6. To facilitate comparisons we reorder *all* results according to the values of the 3rd metric of Multi-GS (the number of minimal subsets sampled before hitting at least one all-inlier minimal subset from each structure) in *ascending order*. This yields an approximate left-to-right ordering based on the difficulty of the different sequences.

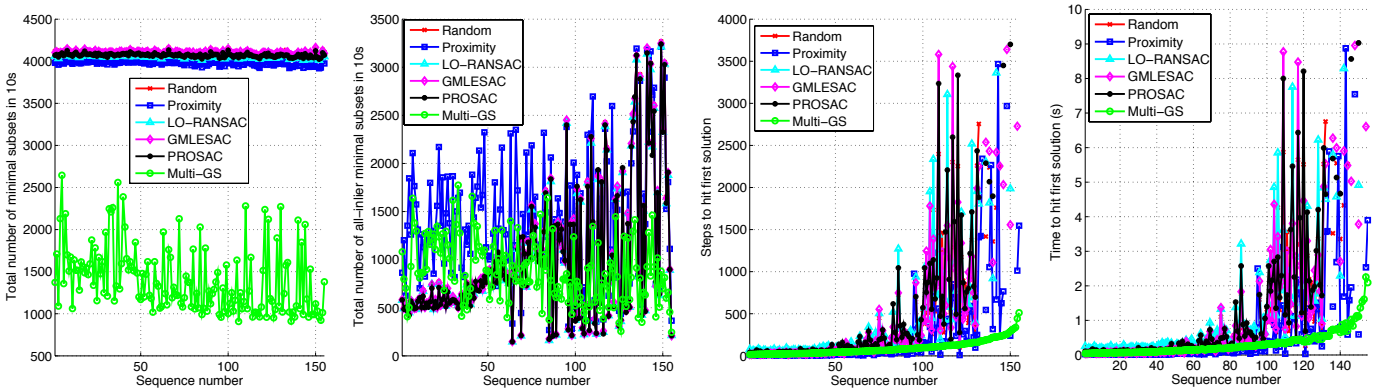
While Fig. 11(b) shows that Multi-GS sampled fewer number of all-inlier minimal subsets, it should be noted that here we are dealing with multi-structure data. The metric used in Fig. 11(b) does not consider from *which* structure the all-inlier minimal subsets were sampled. Due to a lower order geometric model ($p = 4$), with brute speed the other methods simply sampled almost exclusively from the dominant structures (i.e. motions with a large number of trajectories). Hence the other methods actually require more steps/time to hit at least one all-inlier minimal subsets from *each* structure; This is clear from Figs. 11(c) and 11(d). Further, only Multi-GS can successfully hit all structures within the time limit on *all* 155 sequences; The broken lines in the figures indicate failures of the other methods at some of the sequences. Finally note that the divergence in performance between Multi-GS and others occurs close to the 120-th sequence. This is because most of the sequences before number 120



(a) Wadham College (b) Mitchell Building (c) Napier Building

(d) Merton College 1 (e) Merton College 2 (f) Merton College 3

Fig. 10: Image pairs for homography estimation experiment with marked keypoints. The keypoints are detected and matched using SIFT. The colours indicate group labels while gross outliers are marked as '+'. All data are **multi-structure**. (a)(d)(e)(f) are from the Oxford Visual Geometry Group while the others are captured in our campus.



(a) Number of minimal subsets produced within 10s. (b) Number of all-inlier minimal subsets produced within 10s. (c) Number of iterations before hitting at least one all-inlier minimal subset from each structure. (d) Time elapsed before hitting at least one all-inlier minimal subset from each structure.

Fig. 11: Results on the Hopkins 155 dataset (motion subspace estimation) (best viewed in colour).

are easy and we were unable to test at a finer range of difficulty without artificially modifying the data.

Finally since there are no gross outliers in each sequence, all data are *a priori* equally valid inliers. Thus in this experiment, for G-MLESAC and PROSAC we set equal prior inlier probabilities for all data; in any case it is unknown how to compile the KLT tracker results into quality scores analogous to two-view point matching scores. However this also means that the two methods are effectively conducting pure random sampling.

6 Conclusions and Extensions

We propose a fundamentally new accelerated hypothesis sampling scheme which uses information derived from residual sorting. In contrast to existing sampling techniques, our method does not require prior inlier scores from domain knowledge typical in other techniques. The proposed method performs a conditional sampling strategy that encourages selecting minimal subsets within coherent structures. Experiments on various geometric fitting tasks show that the proposed method significantly outperforms other methods in terms of sampling efficiency and speed to recover all-inlier minimal subsets, especially on multi-structure data.

It is evident that Multi-GS excels on both multi- and single-structure data; according to the top row of Fig. 6, Multi-GS is only surpassed by PROSAC on single-structure data. Thus even in cases where the existence of single and multiple structures are possible, using Multi-GS alone will provide competitive results. However, in some applications the practitioner might wish to apply the optimal algorithm in either case. We recommend a solution which involves fusing Multi-GS with PROSAC/Guided-MLESAC. The idea is to sample the first datum s_1 according to prior inlier probabilities such as keypoint matching scores (instead of purely randomly at the moment in Multi-GS), while the other members $s_2, \ldots, s_p$ are sampled according to the Multi-GS conditional weights. This provides the advantage of

TABLE 3: Performance of various sampling methods for multi-homography estimation on image pairs in Fig. 10. Each data was given 50 random runs of 10 CPU seconds each. The same performance metrics as Table 2 are used.

Data		Random	Proximity	LO-RAN SAC	Guided-MLESAC	PROSAC	Multi-GS
Wadham College Fig. 10(a)	Number of samples	**6654**	6268	6580	6580	6588	1723
	Structure-1 (52, 17.7%)	6	17	7	18	30	**50**
	Structure-2 (65, 22.2%)	14	31	24	19	29	**80**
	Sampling Steps	778	338	1080	411	361	**171**
	CPU seconds	1.18	0.55	1.69	0.63	0.55	**0.32**
Mitchell Building Fig. 10(b)	Number of samples	**6580**	6212	6517	6519	6530	1591
	Structure-1 (19, 5.5%)	0	0	0	0	0	0
	Structure-2 (78, 22.6%)	15	27	27	**71**	66	42
	Structure-3 (29, 8.4%)	0	0	0	**1**	0	**1**
	Sampling Steps	×	×	×	×	×	×
	CPU seconds	×	×	×	×	×	×
Napier Building Fig. 10(c)	Number of samples	**6547**	6181	6477	6483	6493	1598
	Structure-1 (76, 20.8%)	11	33	12	24	40	**47**
	Structure-2 (52, 14.2%)	2	5	3	5	7	**10**
	Structure-3 (58, 15.8%)	3	8	4	11	**18**	8
	Sampling Steps	3069	1145	2999	955	677	**449**
	CPU seconds	4.69	1.86	4.64	1.48	**1.05**	1.06
Merton College 1 Fig. 10(d)	Number of samples	**6657**	6271	6589	6586	6592	1862
	Structure-1 (63, 21.3%)	13	33	22	35	48	**53**
	Structure-2 (29, 9.8%)	0	2	0	1	2	**13**
	Structure-3 (8, 2.7%)	0	0	0	0	0	0
	Strucutre-4 (13, 4.4%)	0	0	0	0	0	0
	Sampling Steps	×	×	×	×	×	×
	CPU seconds	×	×	×	×	×	×
Merton College 2 Fig. 10(e)	Number of samples	**6707**	6311	6643	6639	6676	1801
	Structure-1 (41, 15.1%)	3	9	3	8	9	**29**
	Structure-2 (69, 25.5%)	27	56	41	62	82	**109**
	Sampling Steps	1584	483	1587	746	**135**	217
	CPU seconds	2.35	0.77	2.40	1.12	**0.20**	0.41
Merton College 3 Fig. 10(f)	Number of samples	**6672**	6287	6611	6616	6632	1901
	Structure-1 (27, 9.4%)	0	1	0	1	2	**4**
	Structure-2 (30, 10.4%)	1	1	1	1	**3**	**3**
	Structure-3 (22, 7.6%)	0	0	0	0	0	**1**
	Structure-4 (15, 5.2%)	0	0	0	0	0	0
	Sampling Steps	×	×	×	×	×	×
	CPU seconds	×	×	×	×	×	×

the best single-structure algorithms, while retaining the ability of Multi-GS to deal with multiple structures.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their insightful comments. This work was partly supported by the Australian Research Council grant DP0878801.

REFERENCES

[1] D. Capel. An effective bail-out test for RANSAC consensus scoring. In *British Machine Vision Conference (BMVC)*, 2005.
[2] T.-J. Chin, J. Yu, and D. Suter. Accelerated hypothesis generation for multi-structure robust fitting. In *European Conference on Computer Vision (ECCV)*, 2010.
[3] O. Chum and J. Matas. Matching with PROSAC- progressive sample consensus. In *Computer Vision and Pattern Recognition (CVPR)*, 2005.
[4] O. Chum and J. Matas. Optimal randomized RANSAC. *IEEE Trans. on Patt. Analysis and Mach. Intell.*, 30(8):1472–1482, 2008.
[5] O. Chum, J. Matas, and J. Kittler. Locally optimized RANSAC. In *Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM)*, 2003.
[6] O. Chum, T. Werner, and J. Matas. Two-view geometry estimation unaffected by a dominant plane. In *Computer Vision and Pattern Recognition (CVPR)*, 2005.
[7] O. Enqvist and F. Kahl. Two view geometry estimation with outliers. In *British Machine Vision Conference (BMVC)*, 2009.
[8] M. A. Fischler and R. C. Bolles. RANSAC: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24:381–395, 1981.
[9] J.-M. Frahm and M. Pollefeys. RANSAC for (quasi-)degenerate data (QDEGSAC). In *Computer Vision and Pattern Recognition (CVPR)*, 2006.
[10] R. Hartley. In defense of the eight-point algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(6):580–593, 1997.
[11] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2 edition, 2004.
[12] Y. Kanazawa and H. Kawakami. Detection of planar regions with uncalibrated stereo using distributions of feature points. In *British Machine Vision Conference (BMVC)*, 2004.
[13] H. Li. Consensus set maximization with guaranteed global optimality for robust geometry estimation. In *International Conference on Computer Vision (ICCV)*, 2009.
[14] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal on Computer Vision*, 60(2):91–110, 2004.
[15] J. Matas and O. Chum. Randomized RANSAC with $t_{d,d}$ test. *Image and Vision Computing*, 2004.
[16] J. Matas and O. Chum. Randomized RANSAC with sequential probability ratio test. In *Intl. Conf. on Comp. Vis. (ICCV)*, 2005.
[17] K. Ni, H. Jin, and F. Dellaert. GroupSAC: Efficient consensus in the presence of groupings. In *International Conference on Computer Vision (ICCV)*, 2009.
[18] D. Nister. Preemptive RANSAC for live structure and motion estimation. In *Int. Conf. on Computer Vision (ICCV)*, 2003.
[19] R. Raguram, J.-M. Frahm, and M. Pollefeys. A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus. In *European Conference on Computer Vision (ECCV)*, 2008.
[20] P. J. Rousseeuw and A. M. Leroy. *Robust regression and outlier detection*. Wiley, 1987.
[21] T. Sattler, B. Leibe, and L. Kobbelt. SCRAMSAC: Improving RANSAC's efficiency with a spatial consistency filter. In *International Conference on Computer Vision (ICCV)*, 2009.
[22] R. Sedgewick and K. Wayne. *Algorithms*. Addison-Wesley, 4

edition, 2010.
[23] E. Serradell, M. Ozuysal, V. Lepetit, P. Fua, and F. Moreno-Noguer. Combining geometric and appearance priors for robust homography estimation. In *European Conference on Computer Vision (ECCV)*, 2010.
[24] C. V. Stewart. Robust parameter estimation in Computer Vision. *SIAM Review*, 41(3):513–537, 1999.
[25] B. J. Tordoff and D. W. Murray. Guided-MLESAC: Faster image transform estimation by using matching priors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(10):1523–1535, 2005.
[26] R. Tron and R. Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. In *Computer Vision and Pattern Recognition (CVPR)*, 2007.

Tat-Jun Chin received a B.Eng. in Mechatronics Engineering from *Universiti Teknologi Malaysia (UTM)* in 2003, and subsequently in 2007 a Ph.D. in Computer Systems Engineering from Monash University, Victoria, Australia. He was a Research Fellow at the Institute for Infocomm Research (I^2R) in Singapore from 2007–2008. Since 2008 he is a Lecturer at The University of Adelaide, South Australia. His research interests include robust estimation and statistical learning methods in Computer Vision.

Jin Yu received the Master's degree in Artificial Intelligence from the Katholieke Universiteit Leuven, Belgium, in 2004 and the Ph.D. in Information Sciences and Engineering from the Australian National University, Australia, in 2010. She is currently a postdoctoral researcher with the School of Computer Science at The University of Adelaide, Australia. Her main research interests are in stochastic learning, nonsmooth optimisation, and robust model fitting for Machine Learning and Computer Vision problems.

David Suter received a B.Sc.degree in applied mathematics and physics (The Flinders University of South Australia 1977), a Grad. Dip. Comp. (Royal Melbourne Institute of Technology 1984)), and a Ph.D. in computer science (La Trobe University, 1991). He was a Lecturer at La Trobe from 1988 to 1991; and a Senior Lecturer (1992), Associate Professor (2001), and Professor (2006-2008) at Monash University, Melbourne, Australia. Since 2008 he has been a professor in the school of Computer Science, The University of Adelaide. He is head of the School of Computer Science. He served on the Australian Research Council (ARC) College of Experts from 2008-2010. He is on the editorial boards of International Journal of Computer Vision and the Journal of Mathematical Imaging and Vision. He has previously served on the editorial boards of Machine Vision and Applications and the International Journal of Image and Graphics. He was General co-Chair or the Asian Conference on Computer Vision (Melbourne 2002) and is currently co-Chair of the IEEE International Conference on Image Processing (ICIP2013).