IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 19, 2019, accepted January 2, 2020, date of publication January 6, 2020, date of current version January 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2964425

Two-View Geometry Estimation Using RANSAC With Locality Preserving Constraint

GANG WANG, XIAOLIANG SUN, YANG SHANG, ZI WANG, ZHONGCHEN SHI, AND QIFENG YU

College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China
Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation, Changsha 410073, China

Corresponding author: Yang Shang (shangyang1977@nudt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 11727804, Grant 11872070, and Grant 11902349, and in part by the Hunan Provincial Natural Science Foundation of China under Grant 2019JJ50732.

ABSTRACT The random sample consensus (RANSAC) based algorithm is widely used in estimating the two-view geometry from image point correspondences. However, it often becomes extremely slow when the data is contaminated by a large percentage of incorrect matches. To address this problem, the paper proposes a new modification of RANSAC called LP-RANSAC that is robust to varying inlier ratios and achieves large computational savings without deterioration in accuracy. LP-RANSAC integrates the locality preserving constraint into the universal RANSAC framework, which prunes most of the unreliable correspondences before the hypothesize-and-verify loop and guides non-uniform sampling to generate and verify promising models earlier. Unlike other guided sampling strategies, the proposed method is simple to implement and does not require any prior information. Extensive experiments performed on the publicly available datasets reveal that LP-RANSAC can achieve more accurate and stable solutions at much lower computational cost (in milliseconds on standard CPU) than state-of-the-art methods, particularly when handling problems with low inlier ratios.

INDEX TERMS Robust estimation, RANSAC, outlier removal, image matching, two-view geometry.

I. INTRODUCTION

Two-view geometry estimation means fitting a geometric model (e.g., a homography or fundamental matrix) from point correspondences between two images of the same scene. It is one of the most basic tasks in computer vision, which plays a significant role in many applications such as image registration [1]–[3], three-dimensional reconstruction [4], [5] or camera calibration [6]–[8]. The putative point correspondences for two-view geometry estimation typically come from the feature extractors [9]–[11]. As the local descriptor is unlikely to differentiate true and false matches clearly, there are inevitably numerous false matches in the set of data points. This matching problem will be even worse when the image pairs suffer from substantial viewpoint changes, strong occlusion, repeated textures, etc. If the geometric model is directly estimated by some least-squares fitting methods from a set of correspondences that contains even few mismatches, the obtained model may be far away from the true model. Consequently, robust estimation is indispensable for the mismatch removal and estimation of model parameters from perceived data in the presence of measurement noise and outliers (correspondences that are not consistent with the true model).

Due to its simplicity and robustness, the RANSAC [12] algorithm has become the most popular robust estimator to address the outlier removal problem. RANSAC is an iterative approach based on the classic hypothesize and verify paradigm. In brief, standard RANSAC operates by repeatedly generating models estimated from the smallest possible subsets randomly sampled from the set of putative correspondences, and then the model with the highest support is returned after testing each estimated model for support against the entire set of correspondences. In RANSAC the support is the number of inliers, i.e., correspondences within a given error tolerance of the estimated model. The set of obtained inliers is called the consensus set. The hypothesize-and-verify loop is terminated when the probability of finding

The associate editor coordinating the review of this manuscript and approving it for publication was Hugo Proenca.

a consensus set that is compatible with an incorrect model falls below a manually set confidence value or the maximum number of iterations is reached.

The RANSAC algorithm is very easy to implement and has a high success rate to find the correct model under extreme conditions, which lead to its widespread adoption in practice. Moreover, compared to statistical regression methods [13]–[16] (e.g., M-estimators), RANSAC is capable of finding the correct model when the correspondences contain a tremendous percentage of outliers even more than half the data. But, the vanilla RANSAC algorithm has a few drawbacks as well. A prior mathematical model needs to be selected to determine the two-view geometric relations, which may not work for (quasi-)degenerate data that do not provide sufficient constraints to compute the relation uniquely. The cardinality of the consensus set relies too much on the inlier-outlier threshold. If the threshold is set too low, not enough inliers are found. Furthermore, if the threshold is set too high, a number of outliers are mistakenly taken as inliers. In addition, RANSAC tends to severely degrade and be time-consuming as the level of outliers greatly increases. To ensure that the true model is estimated, all putative hypotheses should be generated and testified, which is obviously unrealistic. Therefore, there is always a trade-off between accuracy and efficiency.

In applications, the greatest challenge for the estimation of model parameters in RANSAC stems from outliers. Detecting and eliminating outliers is an intractable yet important problem. A large amount of outliers not only brings about the failure of the convergence to the right solution of the RANSAC algorithm, but also dramatically aggravates the computational burden, which limits its use in real-time tasks. Thus, it is really justifiable and necessary to decrease the number of outliers before performing RANSAC based estimators.

The purpose of this paper is to first briefly review the previous RANSAC-like approaches and then propose an efficient modification of RANSAC for significantly improving the performance of model estimation and outlier removal, especially when the inlier ratio is low. The initial point correspondences for estimating the two-view geometry are obtained relying on feature extractors. The similarity between the descriptors of the detected keypoints from two images is utilized to establish the match candidates. Nevertheless, to safely judge whether the match is correct, only comparing individual feature information is not enough and often results in plentiful false matches. It will bring great difficulties to following robust model estimation, provided that the feature point correspondences that includes false matches are not well pruned. Unfortunately, to the best of our knowledge, no previous work has adequately solved it.

Inspired by Ma et al. [17], we use the locality preserving constraint of those potential matches to remove most of the outliers while barely discarding inliers. Then RANSAC can generate and testify hypotheses on the smaller set of point correspondences with a higher inlier ratio, which can considerably reduce the runtime. Besides, the locality preserving scores can be used to guide sampling matches for generating models. The putative match with the higher score has more chances to be selected. The RANSAC algorithm will converge faster to a correct model by taking advantage of this guided non-uniform sampling strategy instead of sampling uniformly at random. Unlike other guided sampling methods reported in literature, the proposed method does not rely on prior information, meaning that no assumptions are needed to be made about the input data. Experimental results show that the proposed method performs much better than other state-of-the-art methods including RANSAC and its various variants. Since our modification of RANSAC is mainly based on this locality preserving constraint, we name the proposed method LP-RANSAC.

More concretely, the main contributions of this paper can be summarized as follows:

- The locality preserving constraint is integrated into the universal RANSAC framework, resulting in discarding most of the bad correspondences without adding appreciably to the computational cost before performing the RANSAC procedure. The reduced set of more reliable correspondences with a higher inlier ratio is suitable for accelerating hypothesis generation and verification in RANSAC.
- A new non-uniform sampling strategy based on the locality preserving scores is proposed. Compared to other existing guided sampling strategies, it does not require any prior information, such as matching scores.
- The proposed method is applied to both homography and fundamental matrix estimation, which is quantitatively evaluated on a variety of challenging datasets and compared with state-of-the-art methods.

The remainder of this paper is organized as follows. The following section discusses the related work. In Section III, the details of the proposed modification of RANSAC are shown. Section IV provides experimental results and the performance evaluation of the proposed method on different well-known datasets. Finally, the paper is concluded with a summary in Section V.

II. RELATED WORK

In this section, we give an outline of the related work, emphatically discuss RANSAC and its extensions. Since its publication, RANSAC has overtaken the statistical regression algorithms and heuristics as the most widely used and powerful tool for robust estimation in computer vision. In recent forty years, numerous modifications have been proposed aimed at enhancing the performance of the original RANSAC algorithm in different aspects, consisting of accuracy, robustness and efficiency. As previously mentioned, RANSAC works in a hypothesize-and-verify framework. Raguram et al. extend this simple framework to a generalization of RANSAC-like robust estimators, which is termed as a universal RANSAC (USAC [18]) framework, as illustrated in Fig. 1. The USAC framework is mainly composed of

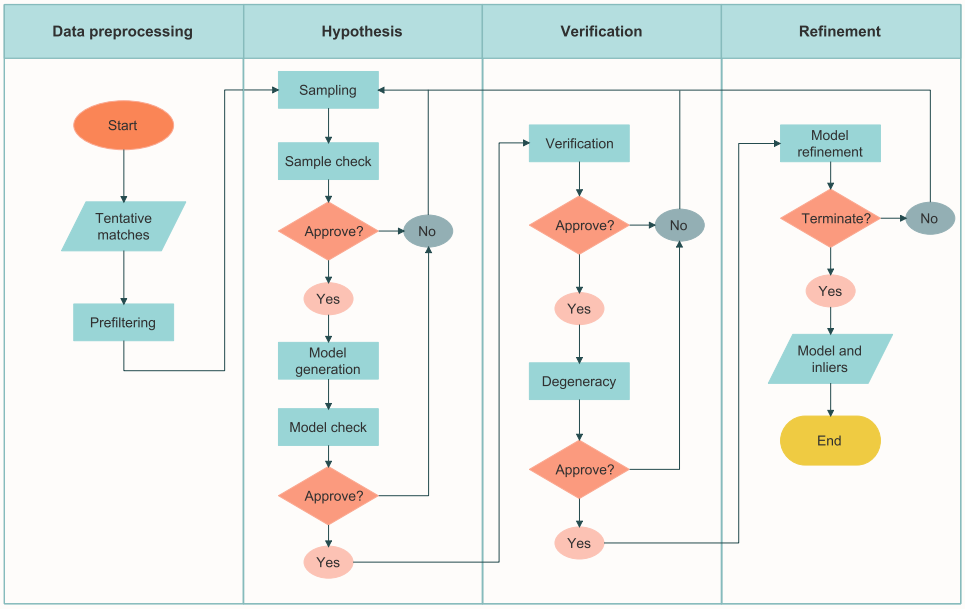


FIGURE 1. The universal RANSAC framework. The framework has the following main steps: data preprocessing, hypothesis generation, hypothesis verification, and model refinement.

data preprocessing, hypothesis, verification and model refinement. RANSAC and a majority of its variants are allowed for the incorporation into the universal framework, which can induct researchers to modify RANSAC specifically and evaluate RANSAC variants systematically. In the following, we survey and discuss most of the important and relevant RANSAC-like methods proposed previously according to the four components of the universal framework.

Considering that the inlier ratio of the correspondence set has a dramatic effect on the runtime of RANSAC, a few efforts have been made to increase the quality of the input data, i.e., detect and remove some outliers. Strictly speaking, this step can be considered as a step of data preprocessing before the RANSAC algorithm commences. Spatially consistent random sample consensus (SCRAMSAC [19]) improves the efficiency of RANSAC by applying a spatial consistency filter to the initial set of point correspondences. This spatial consistency check results in a reduced set of matches with a higher inlier ratio, in turn accelerating the convergence of the remaining iterative process. Similarly, the spatial consistency on RANSAC (SC-RANSAC [20]) algorithm utilizes spatial relations between extracted feature points in two images. A few reliable point correspondences are selected as the base-points to make decisions about the correctness of other matches. This method does not require any prior knowledge about initial data points contrary to SCRAMSAC.

There have been a number of recent efforts that attempt to use guided sampling for minimal set generation rather than uniformly selecting the samples at random, potentially yielding considerable computational savings. The N adjacent points sample consensus (NAPSAC [21]) algorithm assumes that inliers are generally closer to one another than outliers. NAPSAC samples sets of adjacent points lying on a hypersphere of a defined radius under this assumption. The progressive sample consensus (PROSAC [22]) algorithm semi-randomly draws samples from progressively larger sets containing top-ranked matches by a quality measure of tentative correspondences. The group sampling (GroupSAC [23]) algorithm separates data points into several groups where each group has either a high or a low inlier ratio. In GroupSAC, the groups with high inlier ratios have more participation in hypothesis generation based on a binomial mixture model. Another important innovation of the guided sampling algorithm is RANSAC based on extreme value theory (EVSAC [24]). As the name implies, EVSAC models the statistics of the best matching scores and computes a confidence value for each candidate match with extreme value theory. Then the computed confidence values can be applied to form a discrete distribution over the putative correspondences for sampling and hypothesis generation.

Partial evaluation is an alternative solution for speeding up the RANSAC-based model fitting algorithms as well as guided sampling. In partial evaluation, the model estimated from a minimal sample will not be tested on the complete set of data points unless the model passes the test on a small number of data points. Consequently, partial evaluation can filter out bad hypotheses and reduce the time complexity by decreasing the number of data points for hypothesis verification. In randomized RANSAC (R-RANSAC [25]), model verification is first performed on a few randomly selected points. The remaining points are evaluated only if the first selected points are all inliers. This work was further extended by Capel [26] with an effective bail-out test. This bail-out test permits the scoring process for the consensus set to be

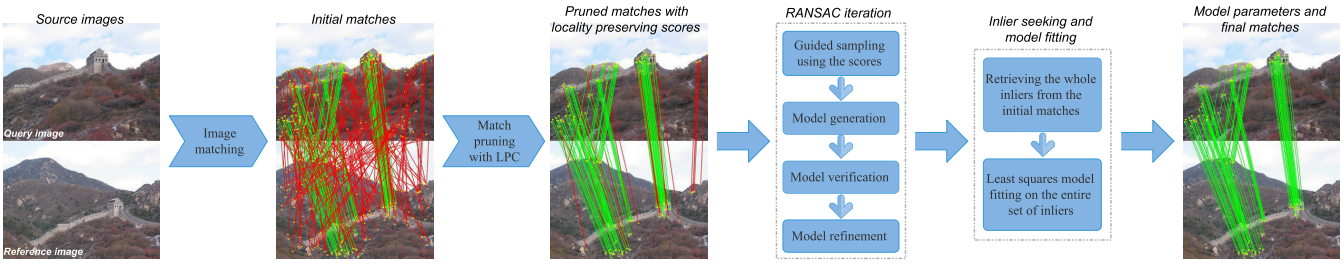


FIGURE 2. Overall flow diagram of the proposed method. Green and red lines represent inliers and outliers respectively.

terminated early. In a general sense, neither of these partial evaluation methods is optimal. Hence, an optimal hypothesis evaluation method based on Wald's theory of sequential decision theory was presented, which is known as WaldSAC [27]. WaldSAC replaces the $T_{d,d}$ test in the hypothesis evaluation stage of R-RANSAC by sequential probability ratio test. How to choose an accurate measurement of model quality is another problem to consider in the verification stage. The maximum likelihood estimation sample consensus (MLESAC [28]) algorithm takes the model support as the log likelihood of the solution instead of just the number of inliers in standard RANSAC. The RANSAC for (quasi-) degenerate data (QDEGSAC [29]) algorithm uses a model selection technique for dealing with degeneracy to estimate the correct model.

In general, the model estimated from the minimal sample is not quite accurate because data points are corrupted with noise. Thus, a model refinement procedure is adopted to achieve acceptable accuracy. The locally optimized RANSAC (LO-RANSAC [30]) algorithm embeds an optimization process into the plain RANSAC using the current best model as a starting point to improve the quality of the solution. The graph-cut RANSAC (GC-RANSAC [31]) algorithm applies the graph cut technique by exploiting the spatial coherence in the local optimization step when a current best model is found.

The RANSAC-like methods typically rely on a provided parametric model, which will fail when the underlying image transformation is unknown in advance. A variety of non-parametric robust feature matching methods have recently been developed to address this problem, including vector field consensus (VFC [32]), learning for mismatch removal (LMR [33]), robust feature matching using spatial clustering method of applications with noise (RFM-SCAN [34]), locality preserving matching (LPM [17]), etc. VFC interpolates a vector field to estimate a consensus of inliers following the non-parametric geometrical constraint. From a novel perspective, LMR and RFM-SCAN respectively cast the feature matching into a two-class classification problem and a spatial clustering problem with outliers by exploiting machine learning techniques. LPM attempts to remove the outliers from given putative point correspondences by preserving the spatial neighborhood relationship among feature points.

LPM has great advantages in terms of speed (more than one order of magnitude faster) in comparison with other non-parametric matching methods, which can accomplish the mismatch removal from over 1000 putative matches in only a few milliseconds. It is beneficial for many real-time tasks and can quickly provide a proper initialization for more sophisticated matching problems. However, LPM primarily concentrates on the outlier identification and removal for robust feature matching, regardless of the transformation model between two images, which is not suitable for model estimation. Since LPM maintains high speed, it can be employed to provide a quick prefiltering for RANSAC to recover the two-view relation.

III. LP-RANSAC

This section provides the proposed LP-RANSAC for fast and robustly estimating the geometric relations between two images. To achieve this goal, we begin with an introduction of the locality preserving constraint (LPC) for filtering out most outliers in the initial set of putative matches. Then this constraint is seamlessly integrated into the universal RANSAC framework, leveraging the score of each correspondence obtained from the locality preserving matching to guide sampling. We finally explain the implementation of the proposed algorithm. The flow diagram of the proposed method is depicted in Fig. 2. In the following, the key steps of the proposed method are described in detail.

A. LOCALITY PRESERVING CONSTRAINT

The initial point correspondences are usually provided by feature-based methods, which involve three main steps: feature detection, feature description, and feature matching. Given an image pair comprised of a query image and a reference image, the feature points on both images can be detected and described using a bunch of famous feature extractors [9]–[11], e.g., the scale invariant feature transform (SIFT [9]). Correspondences between two images are then established by associating each match with a minimum distance between the query descriptor and each of the reference descriptors. By design, these descriptors are normally invariant to scale changes, rotation or illumination, to some extent, affine transformations. However, the above matching process, solely exploiting the feature descriptors described based on the photometric information, will inevitably produce many

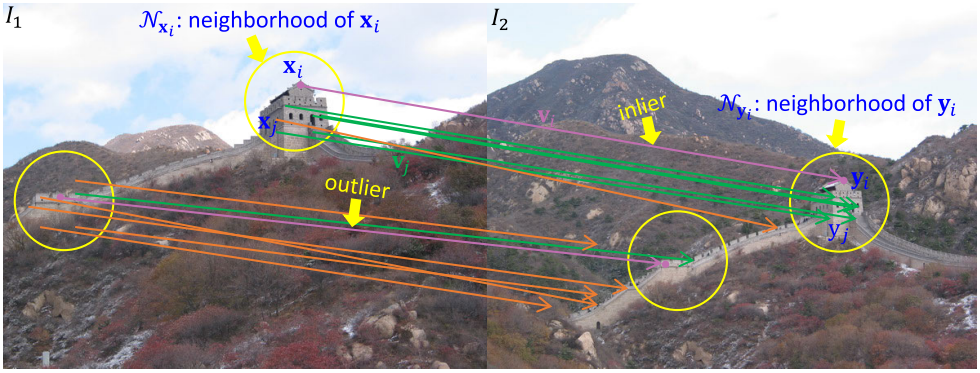


FIGURE 3. Schematic illustration of the locality preserving constraint. An inlier $(\mathbf{x}_i, \mathbf{y}_i)$ will preserve local neighborhood structures (consensus of neighborhood elements and topology), whereas an outlier will not.

false matches. These ambiguities could arise from several factors such as repetitive structures in the scene, image degradation, changes of viewpoint, and others.

It is evident that spatial consistencies and topology relationships among point correspondences can offer stronger constraints to tell correct and incorrect matches apart. Recently, some methods [17], [35]–[38] have been devoted to finding reliable matches by applying this idea to the matching problem. Technically, the locality preserving matching method belongs to this category. LPM is able to yield high quality matches and provide a correctness confidence value for each putative correspondence in a few milliseconds. As a result, it is chosen as a preprocessing step for our LP-RANSAC rather than other similar methods that exploit physical constraints.

In [17], Ma et al. observed that for two images of the same object or scene taken under different viewing conditions, the absolute distance between two feature points may change dramatically, but the relative location among feature points is generally well preserved owing to physical constraints. Based on this observation, a mathematical model for the general feature matching problem was defined, which aimed to constrain the unknown inliers to preserve the local neighborhood structures, as illustrated in Fig. 3.

Suppose that a set of N putative feature correspondences $\mathcal{S} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N}$ has been extracted from a query image I_1 and a reference image I_2, where $\mathbf{x}_i$ and $\mathbf{y}_i$ are 2D vectors denoting the coordinates of feature points. The goal of correspondence pruning is to remove bad correspondences contained in $\mathcal{S}$ to establish reliable matches. To this end, Ma et al. designed an objective function to exploit the consensus of neighborhood elements and topology, which can embrace both rigid and non-rigid deformations, defined as

$$C(\mathbf{l}) = \sum_{i=1}^{N} l_i(c_i - \lambda) + \lambda N, \quad i \in \{1, 2, \cdots, N\}, \tag{1}$$

where λ is a parameter balancing the terms, $\mathbf{l} = [l_1, l_2, \cdots, l_N]$ is a binary vector of N labels $l_i \in \{0, 1\}$ representing the match correctness of the i-th correspondence $(\mathbf{x}_i, \mathbf{y}_i)$, and c_i measures if $(\mathbf{x}_i, \mathbf{y}_i)$ satisfies the geometric constraint of preserving the local neighborhood structure, which can be calculated as

$$c_i = \sum_{m=1}^{M} \frac{1}{MK_m} \left(\sum\nolimits_{j|\mathbf{x}_j \in \mathcal{N}_{\mathbf{x}_i}^{K_m}} d\left(\mathbf{y}_i, \mathbf{y}_j\right) + \sum\nolimits_{j|\mathbf{x}_j \in \mathcal{N}_{\mathbf{x}_i}^{K_m}, \mathbf{y}_j \in \mathcal{N}_{\mathbf{y}_i}^{K_m}} d\left(\mathbf{v}_i, \mathbf{v}_j\right) \right), \tag{2}$$

where d represents a certain distance metric such as Euclidean distance, $\mathbf{v}_i = \overrightarrow{\mathbf{x}_i\mathbf{y}_i}$ denotes a displacement vector, $\mathcal{N}_{\mathbf{x}_i}^{K_m}$ denotes the neighborhood of point $\mathbf{x}_i$ composed of its K_m nearest neighbors, $\mathbf{K} = \{K_m\}_{m=1}^{M}$ is sizes of a set of neighborhoods for the multi-scale neighborhood representation, as demonstrated in Fig. 3. A detailed derivation of (2) can be found in [17], which will not be discussed here. Our focus is to highlight the importance of the locality preserving constraint and to propose a new modification of RANSAC with this constraint.

Since the cost values, or locality preserving scores, $\{c_i\}_{i=1}^{N}$ can be computed in advance, the only unknown variable in (1) is l_i. Clearly, the value of the objective function will decrease if c_i is smaller than λ, whereas it will increase if c_i is larger than λ. Therefore, the optimal solution of $\mathbf{l}$ to this optimization problem is trivially determined by

$$l_i = \begin{cases} 1, & c_i \le \lambda \\ 0, & c_i > \lambda, \end{cases} \quad i \in \{1, 2, \cdots, N\}. \tag{3}$$

After calculating each cost value of the point correspondences, the putative matches can be categorized as inliers and outliers by comparing its corresponding cost value with the predefined value λ according to (3). That is to say, the inlier set is determined depending only on a threshold value of λ. LPM works well as a preprocessor to RANSAC-like methods due to the following reasons. First, the LPM algorithm has linearithmic time complexity, which can satisfy the needs of real-time applications. Specifically, LPM is able to accomplish the outlier removal from over 1000 correspondences in only a few milliseconds. Second, LPM can eliminate most of the outliers and simultaneously preserve most of the inliers,

which considerably increases the inlier rate. Third, LPM can remove false matches near epipolar lines when estimating the fundamental matrix, which greatly improves the precision of model fitting. Therefore, the proposed method uses the matches pruned by LPM and their cost values as new inputs into the following RANSAC framework to improve the efficiency as well as the accuracy of RANSAC.

B. GUIDED SAMPLING WITH LOCALITY PRESERVING SCORES

Vanilla RANSAC generates minimal samples by drawing uniformly at random from the full set of putative correspondences. This implies that no assumptions are made about the putative matches and all point correspondences are treated equally. Model estimation has to be accomplished as quickly and efficiently as possible for many real-world tasks. However, RANSAC becomes exponentially slower employing this uniform sampling strategy as the inlier ratio drops. It is obviously contrary to the primary intention of finding a minimal sample composed of all inliers early in the sampling process of RANSAC. Thus, the guided non-uniform sampling strategy is proposed to bias the selection of correspondences to generate models with more likelihood of being correct preferentially.

The previous guided sampling methods generally modify the sampling strategy by exploiting not only location information but also various types of prior information such as matching scores to speed up RANSAC. For instance, PROSAC sorts the putative matches in descending order with respect to their qualities. However, the reliance on additional matching information restricts the wide application of these methods. Fortunately, as a by-product of match pruning by the LPM algorithm, the locality preserving score can replace the matching score to guide the sampling procedure due to its capability of differentiating outliers and inliers. In addition, the proposed guided sampling strategy requires no extra prior information, leading to its broad application prospects.

Let $\mathcal{S}' = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N'}$ denote a reduced set of N' pruned correspondences with high qualities, and let $\{c_i\}_{i=1}^{N'}$ be the set of locality preserving scores for every correspondence. To some extent, the score c_i indicates that what degree the putative match $(\mathbf{x}_i, \mathbf{y}_i)$ will be an inlier. Therefore, these scores can be used as correctness confidence values to generate a probability distribution for guided non-uniform sampling. The probability of each putative match in the reduced set being drawn is given by the discrete distribution as follows

$$\begin{aligned} p(w_i) &= P(i|w_1, w_2, \cdots, w_{N'}) \\ &= \frac{w_i}{\sum_{j=1}^{N'} w_j}, \quad i \in \{1, 2, \cdots, N'\}, \end{aligned} \tag{4}$$

where

$$w_i = \exp\left(-\frac{{c_i}^2}{2\sigma^2}\right) \quad \text{with } \sigma = \sqrt{\frac{1}{2N'}\sum_{j=1}^{N'} {c_j}^2} \tag{5}$$

is a weight which is assigned to each match pair $(\mathbf{x}_i, \mathbf{y}_i)$ according to its locality preserving score. The higher the weight assigned to the putative match, the greater the probability of this match being selected. These weights can be then applied to randomly sample matches for generating hypotheses by using (4).

Note that when the prior information (e.g., matching scores or geometrical cues) of the putative correspondences is available, it can also attach to the calculation of the weights. Given the set of matching scores produced by some classic feature matching methods, we can similarly assign a weight to each match with respect to its matching scores using (5). In this case, the result of multiplying this weight by the former weight obtained from the locality preserving score will be a new weight to assess the correctness of a correspondence.

The use of locality preserving scores alleviates the dependence on the prior information, resulting in no need for assumptions of the distributions of true and false matches in the putative correspondences. The proposed method, as well as the standard RANSAC, can be applied to many fields because of this characteristic of making no assumptions about the data points. Besides, since locality preserving scores are kind of geometric constraints, the proposed method still performs well in the case of richly textured scenes where correct and incorrect matches are assembled in the local region. In comparison to other guided sampling methods, the proposed one has an advantage that it provides higher robustness to degenerate configuration.

C. ALGORITHM

The structure of the proposed method is simple. To seek a reduced set of correspondences with a higher inlier percentage, we first remove the gross outliers in the set of tentative correspondences using the locality preserving constraint as a preprocessing step. Subsequently, this constraint can be easily incorporated into the RANSAC procedure. We additionally assign a correctness confidence for each remaining correspondence, which is used to guide non-uniform sampling, resulting in promising models being generated and verified earlier.

However, the preprocessing procedure will also lead to a reduction in the number of true matches. The true match that is removed by mistake cannot be retrieved in the subsequent RANSAC procedure without additional remedies. Clearly, missing parts of the true matches will degrade the precision of the geometrical model estimation. To address this issue, a generalization of LPM named GLPM [39] has recently been published. GLPM designs a guided matching strategy based on LPM, using the matching result on a small putative set with a high inlier ratio to guide the matching on the whole putative set. The RANSAC procedure can provide an estimated geometric model, unlike the LPM method, which can categorize the correspondences into inliers and outliers. Thus, after the termination of RANSAC, we first use the current optimal model to seek the true matches in the original putative set $\mathcal{S}$ instead of the reduced set $\mathcal{S}'$. In general,

Algorithm 1 The LP-RANSAC Algorithm

Input: $\mathcal{S}-$ putative set, $\{\mathbf{K}, \lambda\}-$ parameters in (2) and (3), $\epsilon-$ inlier-outlier threshold, $\eta-$ confidence

Output: $\mathcal{I}^*-$ inlier set, θ^*- model parameters

Match pruning:

$\mathcal{S}', \{c_i\}_{i=1}^{N'} \leftarrow$ Prune matches with LPC;

RANSAC using guided sampling:

$w_i \leftarrow$ Calculate weights using (5);

$p(w_i) \leftarrow$ Determine the discrete distribution using (4);

$k_{\max}, k, s^* \leftarrow \infty, 0, 0$;

while $k < k_{\max}$ **do**

  Draw a minimal subset $\mathcal{S}_k \subset \mathcal{S}'$ by the distribution $p(w_i)$, $|\mathcal{S}_k| = m$;

  $\theta_k \leftarrow$ Estimate model parameters using $\mathcal{S}_k$;

  $\mathcal{I}_k, s_k \leftarrow$ Compute the inlier set and the support of θ_k on $\mathcal{S}'$;

  if $s_k > s^*$ **then**

    $\theta^*, s^*, \mathcal{I}^* \leftarrow \theta_k, s_k, \mathcal{I}_k$;

    Update $k_{\max} = \log(1-\eta)/\log(1-\varepsilon^m)$ with $\varepsilon = |\mathcal{I}^*|/N'$;

  end if

  $k = k + 1$;

end while

$\mathcal{I}, s \leftarrow$ Compute the inlier set and support of θ^* on $\mathcal{S}$;

if $s > s^*$ **then**

  $s^*, \mathcal{I}^* \leftarrow s, \mathcal{I}$;

end if

$\theta^* \leftarrow$ Least squares model fitting using $\mathcal{I}^*$.

this new inlier set obtained from $\mathcal{S}$ will cover the missing inliers in the preprocessing procedure. Then, a more precise geometric model is estimated from the whole inliers by implementing the least squares model fitting technique. This post-processing strategy can comparatively increase the inlier number and improve the precision of model estimation in comparison with GLPM, but with much less computational time.

The whole procedure of our LP-RANSAC has been summarized in Algorithm 1. Let $|\cdot|$ denote the cardinality of a set and $k_{\max}$ be the maximum iteration number. Note that the proposed method is very flexible, which can be also combined with other modification strategies of RANSAC such as local optimization.

While the proposed method may sound like a simple idea, it is able to surprisingly improve the time efficiency of the estimated results. There are two main reasons for this outcome. On the one hand, both hypothesis generation and verification are only operated on a far smaller set of putative correspondences with a higher inlier ratio instead of against the full set with a lower inlier ratio. On the other hand, outlier-free minimal samples for estimating promising models can be drawn earlier due to the guided non-uniform sampling. LP-RANSAC is able to find the correct geometric model from the initial set of correspondences without adding extra input information or prior knowledge. It also does not interfere with modifications on other stages of RANSAC, that is to say, LP-RANSAC can be combined with other RANSAC modifications to improve the efficiency or accuracy. Besides, easy implementation is another advantage of the proposed method over other RANSAC variants.

IV. EXPERIMENTAL RESULTS

In this section, the proposed LP-RANSAC is extensively validated on a number of publicly available datasets of real images for estimation of the two-view relations (homography and epipolar geometry) and compared with the following classic and state-of-the-art methods to further demonstrate its superiority:

- RANSAC [12]: The standard RANSAC is considered as a baseline method for comparison.
- PROSAC [22]: It is the most efficient non-uniform sampling method based on known quality function for every point, which is compared with the proposed guided sampling strategy without any prior information.
- GC-RANSAC [31]: It is the best modification of RANSAC in the local optimization stage, which represents the most accurate solution.
- VFC-RANSAC: It is a kind of combination of the outlier removal approach and RANSAC similar to the proposed method. VFC-RANSAC first uses VFC [32] as an alternative to LPM for filtering out outliers in the preprocessing step, and then performs the standard RANSAC procedure. The same post-processing strategy as our LP-RANSAC is also utilized to retrieve the inliers as many as possible.
- RFM-RANSAC: It is analogous to VFC-RANSAC, only replacing VFC with the RFM-SCAN [34] algorithm to remove part of outliers in advance.
- LP-RANSAC*: LP-RANSAC without the prefiltering stage, which just employs the locality preserving score of each correspondence in the putative set derived from the LPC for guided sampling.

All algorithms are implemented in C++ based on the publicly available codes without parallel computing. The termination probability of all RANSAC variants is set to 0.01, whereas the inlier-outlier threshold is set to different values for different cases as will be introduced later. Parameters of other competitors are set to be the default values as suggested in the literature to achieve their best performances. The values of $\mathbf{K}$ and λ in LP-RANSAC are empirically fixed as $\mathbf{K} = [4, 6, 8]$ and $\lambda = 0.9$ for all subsequent experiments. The experiments are performed on a laptop with a 2.6 GHz Intel Core i7 processor and 16 GB memory, which are repeated 500 times to produce statistically meaningful results.

A. DATASETS

To test homography estimation, we adopted VGG [9] (44 pairs), homogr [31] (16 pairs) and EVD [40] (15 pairs) datasets. Each dataset contains image pairs of different sizes with manually annotated ground truth. The VGG dataset

(a) VGG (b) homogr (c) EVD (d) AdelaideRMF (e) kusvod2

FIGURE 4. Examples of test images from the benchmark datasets including (a) VGG, (b) homogr, (c) EVD, (d) AdelaideRMF, and (e) kusvod2.

provides mostly short-baseline stereo images captured under varying imaging conditions, whilst EVD and homogr offer more challenging image pairs for wide-baseline matching. In addition, AdelaideRMF [41] (23 pairs) and kusvod2 [31] (16 pairs) datasets were downloaded for fundamental matrix estimation. The AdelaideRMF dataset consists of image pairs with point correspondences that are manually labeled as outliers or inliers. The kusvod2 dataset provides image pairs with point correspondences and ground truth fundamental matrices. Fig. 4 shows examples of test images in these benchmark datasets.

B. EVALUATION METRICS

To quantitatively evaluate the performance of the proposed method, we use the following metrics:

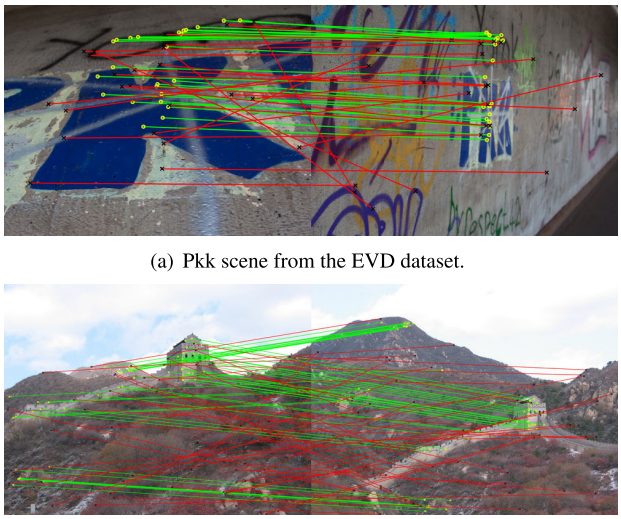
- The processing time t_m is measured in milliseconds (ms).
- The required number of samples N_s and the number of detected inliers $|\mathcal{I}^*|$ are counted.
- The standard deviation of the inliers $\sigma_{\mathcal{I}^*}$ for evaluating the accuracy is defined by

$$\sigma_{\mathcal{I}^*} = \sqrt{\frac{1}{|\mathcal{I}^*|}\sum_{i\in\mathcal{I}^*} e_i^2}, \tag{6}$$

  where e_i^2 is the distance error associated with the i-th inlier. The symmetric transfer error [42] is chosen as the distance measure for homography estimation, and the Sampson error [42] for fundamental matrix estimation. The smaller $\sigma_{\mathcal{I}^*}$ is, the higher the accuracy of the algorithm is.
- The balanced F-score F that combines precision P and recall R is calculated by $F = 2PR/(P+R)$. Precision and recall are defined as

$$P = \frac{N_{tp}}{N_{tp}+N_{fp}}, \quad R = \frac{N_{tp}}{N_{tp}+N_{fn}}, \tag{7}$$

  where N_{tp} denotes the number of true matches, N_{fp} denotes the number of false matches, and N_{fn} denotes the number of matches that are not correctly detected. The bigger F is, the better the algorithm is.



(a) Pkk scene from the EVD dataset.

(b) Wall scene from the kusvod2 dataset.

FIGURE 5. Image pairs chosen to test the robustness of the proposed method for homography (a) and fundamental matrix (b) estimation. For concise exhibition, only a small part of inliers and outliers are plotted. Green and red lines represent inliers and outliers respectively.

C. PERFORMANCE EVALUATION

In general, the inlier ratio is the most important factor that directly determines the efficiency and success of model estimation and outlier removal. Thus, in this section we first test the robustness of the proposed method to putative correspondences with varying inlier ratios for two typical image pairs. Afterward, the performance of the proposed LP-RANSAC tested on the benchmark datasets of more complicated scenarios for the tasks of homography as well as fundamental matrix estimation are demonstrated and compared with that of other competitors. Finally, the limitations of the proposed LP-RANSAC are discussed.

1) ROBUSTNESS TEST

As discussed previously, the performance of two-view geometry estimation strongly depends on the inlier ratio of point correspondences. To evaluate the robustness of the proposed method over varying inlier ratios, we apply it to two

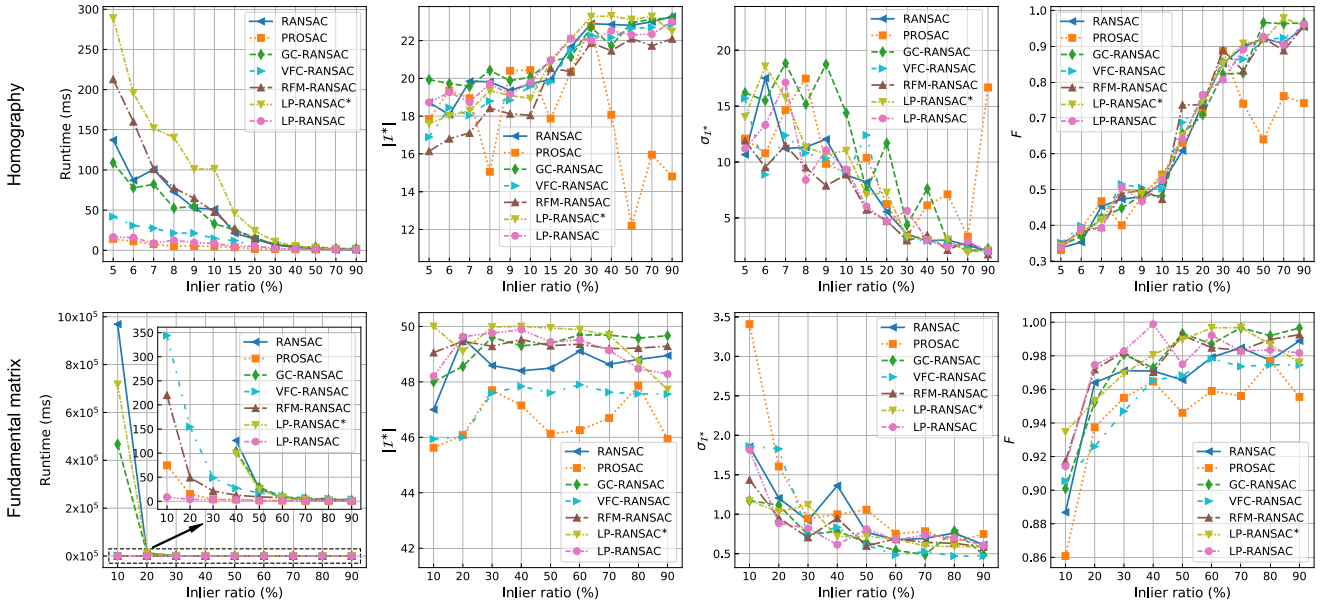


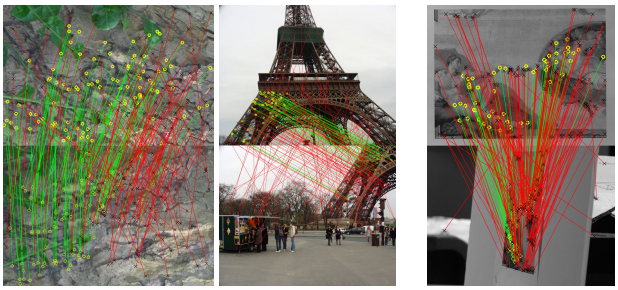
FIGURE 6. Performance of model estimation for homography (top row) and fundamental matrix (bottom row) w.r.t. varying inlier ratios. The runtime, the number of inliers found $|\mathcal{I}^*|$, the standard deviation of the inliers $\sigma_{\mathcal{I}^*}$, and the F-score F for each algorithm are shown from left to right.

typical image pairs (see Fig. 5) chosen from the datasets. For homography and fundamental matrix estimation, we test on the pkk scene from the EVD dataset and the wall scene from the kusvod2 dataset, respectively. The inlier ratios of these two sets of correspondences are varied by adjusting the number of outliers, whereas the number of inliers is fixed. The inlier ratio is gradually increased from 0.05 to 0.9 with a constant inlier number of 24 for the pkk scene, whereas the inlier number is fixed to 50 with the inlier ratio ranged from 0.1 to 0.9 for the wall scene. In both cases, we use evaluation metrics as defined above to assess the robustness of the proposed method and compare the results to those of the aforementioned six representative methods, including three RANSAC variants (RANSAC, PROSAC, and GC-RANSAC) and three combined methods based on the outlier removal and RANSAC (VFC-RANSAC, RFM-RANSAC, and LP-RANSAC*).

The top and bottom rows in Fig. 6 show the results of this experiment for homography and fundamental matrix estimation, respectively. From left to right, the plots measure runtimes, the number of inliers, the standard deviation of inliers and the balanced F-score of LP-RANSAC and the competitive methods.

To explore the role that the locality preserving constraint plays on the proposed LP-RANSAC, we compare our proposed LP-RANSAC with standard RANSAC and LP-RANSAC*. From the results, we see that LP-RANSAC* is faster than RANSAC on fundamental matrix estimation but slower on homography estimation. This is probably because the consuming time the preprocessing step takes is relatively considerable when the estimated model is simple. In view of the number of inliers, the standard deviation of inliers and the F-score, LP-RANSAC and LP-RANSAC* achieve comparable and satisfying performance, but marginally better than RANSAC. This can be explained by the guided sampling with the LPC. These results indicate that the improvement of efficiency is mainly due to the match pruning using the LPC, and the improvement of precision probably arises from the guided sampling according to locality preserving scores. In conclusion, the use of the LPC in both the preprocessing procedure and the RANSAC procedure contributes to the robustness of our LP-RANSAC on different degrees of outlier ratios.

As Fig. 6 shows, the performance of all the seven methods degrades with the decrease of the inlier ratio. It is clear that, in terms of runtime, LP-RANSAC performs equivalent or even marginally faster than PROSAC, while only consuming a fraction of runtime of both RANSAC and GC-RANSAC. Among all combinations of outlier removal methods and RANSAC, LP-RANSAC still has the best efficiency. When the inlier ratio of the putative correspondence set is low, the effect of reducing computational overhead by using LP-RANSAC is more significant than the cases of high inlier ratios. In terms of solution quality, LP-RANSAC remains robust up to large changes of inlier ratios and achieves a comparable result as RANSAC and GC-RANSAC, whereas significantly outperforming PROSAC. Note from the graphs that the number of inliers, the standard deviation of inliers and the balanced F-score of LP-RANSAC are similar to those of RANSAC and GC-RANSAC. Thus the obtained solutions are of the same quality. Compared to PROSAC, it can also be seen that LP-RANSAC yields more accurate and stable results at a comparable runtime in view of the above three metrics. This is due to the fact that the locality



(a) bark (b) Eiffel (c) adam

FIGURE 7. Sample qualitative results of LP-RANSAC for homography estimation tested on (a) EVD, (b) homogr, and (c) EVD datasets. For concise exhibition, only parts of inliers and outliers are plotted. Green and red lines represent detected inliers and outliers respectively.

preserving constraint used in LP-RANSAC can filter out lots of outliers, thus resulting in faster convergence to the right solution. Note that LP-RANSAC performs slightly better than VFC-RANSAC and RFM-RANSAC in terms of precision. The reason is that VFC-RANSAC and RFM-RANSAC falsely remove a large portion of true matches in the preprocessing stage and lack the guided non-uniform sampling in the RANSAC procedure used in our LP-RANSAC. Therefore, taking both runtime and solution quality into consideration, LP-RANSAC has stronger robustness to varying inlier ratios than the competitive estimators.

2) HOMOGRAPHY RESULTS

In this experiment, we assess the performance of LP-RANSAC for the problem of homography estimation by using VGG, homogr and EVD datasets. These datasets have different levels of inlier ratios and deformation, making the estimation of homographies challenging. A putative correspondence is considered as an inlier if the symmetric transfer error of the homography is less than 3 pixels for VGG and homogr datasets, whilst 7 pixels for EVD dataset since each mage pair in this dataset has a very large angular difference between views.

Fig. 7 displays sample qualitative results of LP-RANSAC for homography estimation. The quantitative results of this experiment are summarized in Table 1, reporting the quality of homography estimation on each dataset. The table lists the runtime t_m, the required number of samples N_s, the number of detected inliers $|\mathcal{I}^*|$, the standard deviation of the inliers $\sigma_{\mathcal{I}^*}$, and the F-score F for each method. The values are averaged over 500 executions of each estimator on all tested image pairs of each dataset. The best results are highlighted in bold.

It can be seen that LP-RANSAC has the lowest runtimes and generates fewest samples for most image pairs. This is due to the outlier removal and non-uniform sampling in the algorithm. Note that PROSAC is unexpectedly slower than RANSAC for the VGG dataset, as most image pairs in this dataset have many repetitive structures, resulting in PROSAC struggling to converge to the correct model. LP-RANSAC delivers solutions that are almost the most accurate in terms of the standard deviation of inliers, which capture the

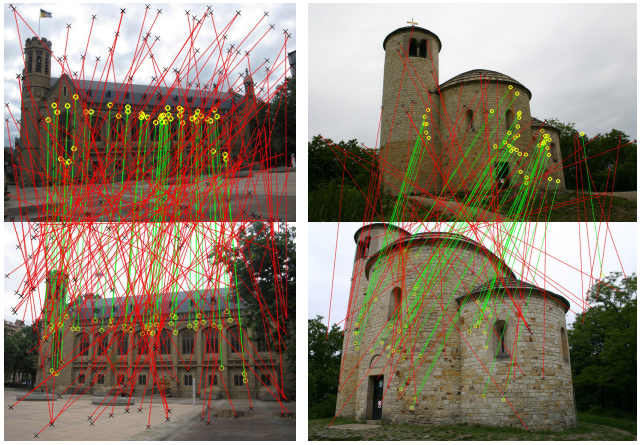
TABLE 1. Comparison of various robust estimators for homography estimation on real-world datasets.

Methods	Metrics	VGG	homogr	EVD
RANSAC	t_m	28.238	37.301	428.298
	N_s	868.598	244.650	20782.791
	$\|\mathcal{I}^*\|$	885.175	251.758	31.245
	$\sigma_{\mathcal{I}^*}$	1.704	2.478	7.360
	F	0.809	0.629	0.477
PROSAC	t_m	45.125	**14.994**	21.673
	N_s	2334.379	**50.442**	675.427
	$\|\mathcal{I}^*\|$	694.876	227.161	29.777
	$\sigma_{\mathcal{I}^*}$	1.748	**2.438**	6.912
	F	0.740	0.600	0.474
GC-RANSAC	t_m	61.826	52.886	370.114
	N_s	831.439	166.725	18252.332
	$\|\mathcal{I}^*\|$	**918.680**	**265.843**	**33.590**
	$\sigma_{\mathcal{I}^*}$	1.725	2.450	7.341
	F	**0.821**	**0.645**	0.480
VFC-RANSAC	t_m	38.946	43.160	90.714
	N_s	639.091	108.385	3330.145
	$\|\mathcal{I}^*\|$	881.166	251.929	29.804
	$\sigma_{\mathcal{I}^*}$	3.245	2.476	7.928
	F	0.803	0.630	0.468
RFM-RANSAC	t_m	2979.386	1887.226	295.299
	N_s	**24.065**	108.949	960.108
	$\|\mathcal{I}^*\|$	881.582	249.019	30.124
	$\sigma_{\mathcal{I}^*}$	1.737	2.477	7.138
	F	0.802	0.619	0.470
LP-RANSAC*	t_m	45.921	58.185	405.393
	N_s	706.640	196.544	17824.948
	$\|\mathcal{I}^*\|$	898.428	260.266	32.600
	$\sigma_{\mathcal{I}^*}$	1.717	2.477	7.151
	F	0.818	0.635	0.484
LP-RANSAC	t_m	**22.300**	31.273	**16.379**
	N_s	59.039	101.411	**617.053**
	$\|\mathcal{I}^*\|$	885.804	253.208	31.414
	$\sigma_{\mathcal{I}^*}$	**1.689**	2.509	**6.654**
	F	0.810	0.628	**0.485**

second most inliers after GC-RANSAC. This is because the local optimization technique exploited in GC-RANSAC is not incorporated in LP-RANSAC on consideration of runtime. For the same reason, LP-RANSAC reaches approximately similar results as RANSAC in terms of the balanced F-score, slightly inferior to GC-RANSAC. VFC-RANSAC and RFM-RANSAC do not achieve satisfying performance in speed as LP-RANSAC, because outlier removal in both two methods takes too much time, especially when the number of putative feature matches is large. Compared to full LP-RANSAC, the performance of LP-RANSAC* is comparable but many orders of magnitude slower in the cases where the inlier ratio in the putative set is very small. This result gives a further verification on the idea that the reduced set with a higher inlier ratio can significantly speed up RANSAC. Given the above discussion, LP-RANSAC is thus able to return more accurate and stable results than the competitive methods at an affordable cost, in particular when the inlier ratio is low.

3) FUNDAMENTAL MATRIX RESULTS

In this experiment, we evaluate the performance of LP-RANSAC in estimating fundamental matrices by adopting AdelaideRMF and kusvod2 datasets. These datasets cover



(a) bonython (b) rotunda

FIGURE 8. Sample qualitative results of LP-RANSAC for fundamental matrix estimation tested on (a) AdelaideRMF and (b) kusvod2 datasets. For concise exhibition, only parts of inliers and outliers are plotted. Green and red lines represent detected inliers and outliers respectively.

a wide range of challenging scenarios, where the inlier ratios ranged from 23%–94%. Correspondences are considered to be inliers if the Sampson error of the fundamental matrix is smaller than 3 pixels.

Sample qualitative results of LP-RANSAC are shown in Fig. 8, and the averaged results calculated using all the image pairs of the two datasets are shown in Table 2. The same statistics are presented as in the homography experiment. As before, it can be clearly seen that LP-RANSAC consistently obtains accurate and stable results with a low computational cost. LP-RANSAC tends to require in general much fewer samples and lower runtimes than the other methods to generate a correct model, save for PROSAC, which can sometimes be slightly faster than LP-RANSAC. However, compared to standard techniques, PROSAC typically delivers inaccurate solutions with much fewer inliers, larger standard deviations of inliers and smaller balanced F-score, whereas the corresponding values for LP-RANSAC are comparable or even preferable as RANSAC and GC-RANSAC. This again indicates that the proposed method is superior to the other competitive RANSAC-like methods. Similar to the homography case, LP-RANSAC's performance is much higher than other combined methods like VFC-RANSAC and RFM-RANSAC in both speed and precision, and is comparable to LP-RANSAC* which is many orders of magnitude slower. Interestingly, LP-RANSAC* returns more inliers, lower standard deviations of inliers and higher F-score than full LP-RANSAC on the kusvod2 dataset. This is due to that LP-RANSAC improperly removes a portion of inliers, which may not be successfully retrieved by the post-processing strategy, leading to an inaccurate solution to the fundamental matrix estimation.

4) LIMITATIONS

Because of the nature of RANSAC, our LP-RANSAC will also suffer from the problem of the inlier-outlier threshold setting as other RANSAC variants, which strongly influences the outcome of feature matching or model estimation. In addition, the outlier removal as a preprocessing filter for the subsequent RANSAC will not always bring benefit to the performance of the proposed method. LP-RANSAC might fall behind the standard RANSAC in speed when the putative correspondence set has sufficient inliers and a high inlier ratio, as the prefiltering unnecessarily takes some time in this case. Another potential limitation of our LP-RANSAC is that the LPC does not guarantee only outliers are eliminated but probably also eliminate inliers. This may be problematic

TABLE 2. Comparison of various robust estimators for fundamental matrix estimation on real-world datasets.

Methods	Metrics	AdelaideRMF	kusvod2
RANSAC	t_m	169.351	34.989
	N_s	5451.532	925.626
	$\lvert\mathcal{I}^*\rvert$	248.632	313.263
	$\sigma_{\mathcal{I}^*}$	5.800	4.000
	F	0.965	0.891
PROSAC	t_m	**3.669**	6.860
	N_s	77.506	123.283
	$\lvert\mathcal{I}^*\rvert$	224.197	290.205
	$\sigma_{\mathcal{I}^*}$	6.537	4.536
	F	0.929	0.856
GC-RANSAC	t_m	153.797	35.075
	N_s	5152.401	758.845
	$\lvert\mathcal{I}^*\rvert$	**252.004**	319.308
	$\sigma_{\mathcal{I}^*}$	5.510	4.310
	F	0.968	0.897
VFC-RANSAC	t_m	11.202	12.873
	N_s	147.672	80.865
	$\lvert\mathcal{I}^*\rvert$	246.505	303.157
	$\sigma_{\mathcal{I}^*}$	4.860	3.886
	F	0.965	0.870
RFM-RANSAC	t_m	274.257	551.874
	N_s	**7.019**	25.128
	$\lvert\mathcal{I}^*\rvert$	246.632	306.175
	$\sigma_{\mathcal{I}^*}$	4.746	4.025
	F	0.968	0.871
LP-RANSAC*	t_m	168.450	40.448
	N_s	4563.900	699.883
	$\lvert\mathcal{I}^*\rvert$	250.781	**319.357**
	$\sigma_{\mathcal{I}^*}$	4.917	**3.092**
	F	0.968	**0.903**
LP-RANSAC	t_m	3.732	**5.605**
	N_s	19.136	**24.241**
	$\lvert\mathcal{I}^*\rvert$	248.060	312.955
	$\sigma_{\mathcal{I}^*}$	**4.588**	3.261
	F	**0.971**	0.892

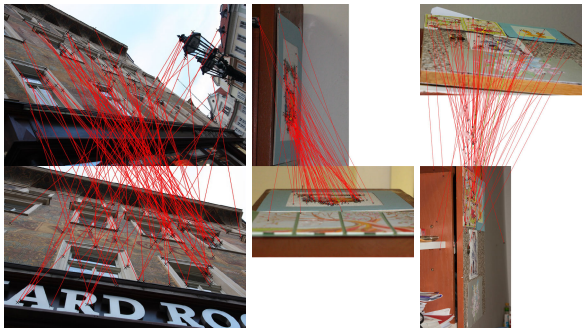


FIGURE 9. Example of failure cases of LP-RANSAC. For concise exhibition, only 100 putative matches are plotted. As can be seen, there are barely any inliers that are correctly detected.

when both the number of inliers and the inlier ratio in the putative set are quite small, leading to the convergence to the wrong solution in the RANSAC procedure. Such a case is most likely to occur with extreme viewpoint changes. However, the performance of LP-RANSAC is still satisfactory for most cases. Typical failure cases are presented in Fig. 9.

V. CONCLUSION

In this paper, we have presented a very fast and accurate robust estimator named LP-RANSAC for recovering the two-view geometry. Our central idea is to integrate the locality preserving constraint within the framework of hypothesize-and-verify without any prior information. By exploiting this constraint, most of the bad matches are pruned and the corresponding locality preserving scores of the remaining matches are obtained in the preprocessing step. Furthermore, these scores are used to guide the non-uniform sampling on the smaller set of more reliable correspondences with a higher fraction of inliers, resulting in seeking the correct model earlier without any deterioration in accuracy. As shown in our experiments, the proposed method is able to provide significantly better solutions to both the homography and fundamental matrix estimation problems in almost all cases than the state-of-the-art methods. Our future work includes combining other modification strategies for RANSAC with our method and extending LP-RANSAC to other applications in computer vision, such as pose estimation, motion segmentation or multi-model fitting.

REFERENCES

[1] M. Lhuillier and L. Quan, "A quasi-dense approach to surface reconstruction from uncalibrated images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 418–433, Mar. 2005.

[2] C. Wu, B. Clipp, X. Li, J.-M. Frahm, and M. Pollefeys, "3D model matching with viewpoint-invariant patches (VIP)," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1229–1236.

[3] Y. Qian, M. Gong, and Y.-H. Yang, "Stereo-based 3D reconstruction of dynamic fluid surfaces by global optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6650–6659.

[4] D. Thomas and A. Sugimoto, "Range image registration using a photometric metric under unknown lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2252–2269, Sep. 2013.

[5] Y. Wu, W. Ma, M. Gong, L. Su, and L. Jiao, "A novel point-matching algorithm based on fast sample consensus for image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 43–47, Jan. 2015.

[6] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artif. Intell.*, vol. 78, nos. 1–2, pp. 87–119, Oct. 1995.

[7] B. Guan, Y. Shang, and Q. Yu, "Planar self-calibration for stereo cameras with radial distortion," *Appl. Opt.*, vol. 56, no. 33, p. 9257, Nov. 2017.

[8] B. Guan, Y. Yu, A. Su, Y. Shang, and Q. Yu, "Self-calibration approach to stereo cameras with radial distortion based on epipolar constraint," *Appl. Opt.*, vol. 58, no. 31, p. 8511, Nov. 2019.

[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[10] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, nos. 1–2, pp. 43–72, Nov. 2005.

[11] L. Zheng, Y. Yang, and Q. Tian, "SIFT meets CNN: A decade survey of instance retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1224–1244, May 2018.

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

[13] P. J. Huber, *Robust Statistics*. Hoboken, NJ, USA: Wiley, 1981.

[14] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*. Hoboken, NJ, USA: Wiley, 1987.

[15] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *Int. J. Comput. Vis.*, vol. 27, no. 2, pp. 161–195, 1998.

[16] C. V. Stewart, "Robust parameter estimation in computer vision," *SIAM Rev.*, vol. 41, no. 3, pp. 513–537, Jan. 1999.

[17] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, May 2019.

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

[19] T. Sattler, B. Leibe, and L. Kobbelt, "SCRAMSAC: Improving RANSAC's efficiency with a spatial consistency filter," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep. 2009, pp. 2090–2097.

[20] M. Fotouhi, H. Hekmatian, M. A. Kashani-Nezhad, and S. Kasaei, "SC-RANSAC: Spatial consistency on RANSAC," *Multimed Tools Appl.*, vol. 78, no. 7, pp. 9429–9461, Apr. 2019.

[21] D. Myatt, P. Torr, S. Nasuto, J. Bishop, and R. Craddock, "NAPSAC: High Noise, High Dimensional Robust Estimation–it's in the Bag," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Cardiff, Wales, 2002, pp. 458–467.

[22] O. Chum and J. Matas, "Matching with PROSAC—Progressive sample consensus," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, Jul. 2005, pp. 220–226.

[23] K. Ni, H. Jin, and F. Dellaert, "GroupSAC: Efficient consensus in the presence of groupings," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep. 2009, pp. 2193–2200.

[24] V. Fragoso, P. Sen, S. Rodriguez, and M. Turk, "EVSAC: Accelerating hypotheses generation by modeling matching scores with extreme value theory," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sydney, NSW, Australia, Dec. 2013, pp. 2472–2479.

[25] J. Matas and O. Chum, "Randomized RANSAC with Td,d test," *Image Vis. Comput.*, vol. 22, no. 10, pp. 837–842, Sep. 2004.

[26] D. P. Capel, "An effective bail-out test for RANSAC consensus scoring," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Oxford, U.K., 2005, p. 78.1.

[27] J. Matas and O. Chum, "Randomized RANSAC with sequential probability ratio test," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, Beijing, China, vol. 1, 2005, pp. 1727–1732.

[28] P. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understand.*, vol. 78, no. 1, pp. 138–156, Apr. 2000.

[29] J.-M. Frahm and M. Pollefeys, "RANSAC for (quasi-)degenerate data (QDEGSAC)," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New York, NY, USA, Jul. 2006, pp. 453–460.

[30] O. Chum, J. Matas, and J. Kittler, "Locally optimized RANSAC," in *Proc. Joint Pattern Recognit. Symp.*, 2003, pp. 236–243.

[31] D. Barath and J. Matas, "Graph-cut RANSAC," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 6733–6741.

[32] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, Apr. 2014.

[33] J. Ma, X. Jiang, J. Jiang, J. Zhao, and X. Guo, "LMR: Learning a two-class classifier for mismatch removal," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4045–4059, Aug. 2019.

[34] X. Jiang, J. Ma, J. Jiang, and X. Guo, "Robust feature matching using spatial clustering with heavy outliers," *IEEE Trans. Image Process.*, vol. 29, no. 29, pp. 736–746, Aug. 2020.

[35] W.-Y.-D. Lin, M.-M. Cheng, J. Lu, H. Yang, M. N. Do, and P. Torr, "Bilateral functions for global motion modeling," in *Proc. ECCV*, Zürich, Switzerland, 2014, pp. 341–356.

[36] W.-Y. Lin, S. Liu, N. Jiang, M. N. Do, P. Tan, and J. Lu, "RepMatch: Robust feature matching and pose for reconstructing modern cities," in *Proc. ECCV*, Amsterdam, The Netherlands, 2016, pp. 562–579.

[37] W.-Y. Lin, F. Wang, M.-M. Cheng, S.-K. Yeung, P. H. Torr, M. N. Do, and J. Lu, "CODE: Coherence based decision boundaries for feature correspondence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 34–47, Jan. 2018.

[38] J. Bian, W.-Y. Lin, Y. Matsushita, S.-K. Yeung, T.-D. Nguyen, and M.-M. Cheng, "GMS: Grid-based motion statistics for fast, ultra-robust feature correspondence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2828–2837.
[39] J. Ma, J. Jiang, H. Zhou, J. Zhao, and X. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4435–4447, Aug. 2018.
[40] D. Mishkin, J. Matas, and M. Perdoch, "MODS: Fast and robust method for two-view matching," *Comput. Vis. Image Understand.*, vol. 141, pp. 81–93, Dec. 2015.
[41] H. S. Wong, T.-J. Chin, J. Yu, and D. Suter, "Dynamic and hierarchical multi-structure geometric model fitting," in *Proc. Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 1044–1051.
[42] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.

GANG WANG received the B.E. degree from Wuhan University, Wuhan, China, in 2013, and the M.S. degree from the National University of Defense Technology, Changsha, China, in 2015, where he is currently pursuing the Ph.D. degree with the College of Aerospace Science and Engineering. His research interests include image matching, pose estimation, and computer vision.

XIAOLIANG SUN received the B.E., M.S., and Ph.D. degrees from the National University of Defense Technology, Changsha, China, in 2010, 2013, and 2017, respectively. He is currently a Lecturer with the College of Aerospace Science and Engineering, National University of Defense Technology. His main research interests include image measurement, machine vision, and navigation systems.

YANG SHANG received the B.E., M.S., and Ph.D. degrees in aerospace science and technology from the National University of Defense Technology, Changsha, China, in 1998, 2000, and 2005, respectively. He is currently a Professor with the College of Aerospace Science and Engineering, National University of Defense Technology. His main research interests include photogrammetry, machine vision, and navigation systems.

ZI WANG received the B.E. degree from Tianjin University, Tianjin, China, in 2016, and the M.S. degree from the National University of Defense Technology, Changsha, China, in 2018, where he is currently pursuing the Ph.D. degree with the College of Aerospace Science and Engineering. His research interests include deep learning, object pose detection, and computer vision.

ZHONGCHEN SHI received the B.E. degree from Beihang University, Beijing, China, in 2014, and the M.S. degree from the National University of Defense Technology, Changsha, China, in 2016, where he is currently pursuing the Ph.D. degree with the College of Aerospace Science and Engineering. His research interests include 3D reconstruction, object pose estimation, and tracking.

QIFENG YU received the B.S. degree from Northwestern Polytechnic University, Xi'an, China, in 1981, the M.S. degree from the National University of Defense Technology, Changsha, China, in 1984, and the Ph.D. degree from Bremen University, Bremen, Germany, in 1996.

He is currently a Professor with the National University of Defense Technology. He has authored three books and published over 100 articles. His main research fields are image measurement, vision navigation, and close-range photogrammetry.

Dr. Yu is a member of the Chinese Academy of Sciences.