Efficient Deterministic Search with Robust Loss Functions for Geometric Model Fitting

Aoxiang Fan, Jiayi Ma, Xingyu Jiang, and Haibin Ling

Abstract—Geometric model fitting is a fundamental task in computer vision, which serves as the pre-requisite of many downstream applications. While the problem has a simple intrinsic structure where the solution can be parameterized within a few degrees of freedom, the ubiquitously existing outliers are the main challenge. In previous studies, random sampling techniques have been established as the practical choice, since optimization-based methods are usually too time-demanding. This prospective study is intended to design efficient algorithms that benefit from a general optimization-based view. In particular, two important types of loss functions are discussed, *i.e.* truncated and l_1 losses, and efficient solvers have been derived for both upon specific approximations. Based on this philosophy, a class of algorithms are introduced to perform deterministic search for the inliers or geometric model. Recommendations are made based on theoretical and experimental analyses. Compared with the existing solutions, the proposed methods are both simple in computation and robust to outliers. Extensive experiments are conducted on publicly available datasets for geometric estimation, which demonstrate the superiority of our methods compared with the state-of-the-art ones. Additionally, we apply our method to the recent benchmark for wide-baseline stereo evaluation, leading to a significant improvement of performance.

Index Terms—Geometric model fitting, robust loss function, deterministic search, outlier, image matching.

1 Introduction

In computer vision, a vast majority of applications, such as structure-from-motion [1], simultaneous localization and mapping [2] and image mosaic [3], rely on feature point correspondences between 2D images to infer the spatial transformation or the 3D geometry [4]. As established by existing research, the solution can be parameterized by different geometric models, such as affine, homography or fundamental matrix, with only several degrees of freedom [5]. However, due to the imperfections of both local key point detection and feature description techniques, the correspondences are invariably contaminated by noise and a number of outliers. The degenerated data pose a great challenge for accurate estimation of the geometric models.

To tackle the problem, the most well-known and widely used method is probably the RANdom SAmple Consensus (RANSAC) algorithm [6], despite its simplicity and age of invention. In essence, RANSAC proceeds by repeatedly sampling a random minimal subset of correspondences to propose hypotheses, *e.g.* 4 correspondences for homography and 7 for fundamental estimation. The process is iterated until a convergence criterion, which provides a probabilistic guarantee of hitting an all-inlier subset, is met. The success of RANSAC is largely attributed to the low degree of freedom of geometric models, for which the random sampling strategy can be applied without excessive computations. However, some fundamental shortcomings exist with the randomized hypothesize-and-verify search strategy. One of the main limitations lies in the degraded performance due to dominant outliers. The required time to retrieve an all-inlier subset grows exponentially with respect to the outlier rate, and the estimation accuracy also suffers from high uncertainty. Although a large amount of literature has been published in the last two decades to improve the primitive RANSAC algorithm toward better efficiency and accuracy [7], [8], the performance is still restricted due to these limitations.

Different from the random sampling techniques, another line of work has focused on optimization-based frameworks to perform deterministic search for geometric model fitting. This perspective is intriguing in that theoretical guarantees can be derived regarding the optimality of the solution [9], [10], [11], [12]. However, the fundamental intractability of the problem means that the globally optimal algorithms must rely on exhaustive search in nature. Consequently, they are only suitable for a small number of correspondences. More recently, there has been an increasing attention given to approximate solutions with locally convergent algorithms [13], [14]. The practicality of these methods has been significantly improved compared to the globally optimal ones. However, albeit alleviated, the issue of high computational expense is still unresolved to meet the requirements for real-world applications, let alone their sensitivity to local optima.

The specific objective of this paper is to investigate efficient deterministic search algorithms for geometric model fitting, which has drawn rather limited attention in the literature. Here by efficiency we mean that the deterministic search stage can be accomplished within tens of milliseconds, allowing the application to real-time tasks. Our method is based on the investigation of robust loss functions, which are of a great variety [14], [15]. In general, there are two properties that are critical to the effectiveness of a robust function, *i.e.* exactness and convexity. Thus although

- *A. Fan, J. Ma and X. Jiang are with the Electronic Information School, Wuhan University, Wuhan, 430072, China (email: fanaoxiang@whu.edu.cn, jyma2010@gmail.com, jiangx.y@whu.edu.cn).*
- *H. Ling is with the Department of Computer Science, Stony Brook University, NY, 11794, USA. (email: haibin.ling@stonybrook.edu).*

Manuscript received Dec. 8, 2020; revised Jun. 2, 2021. (Corresponding author: Jiayi Ma.)

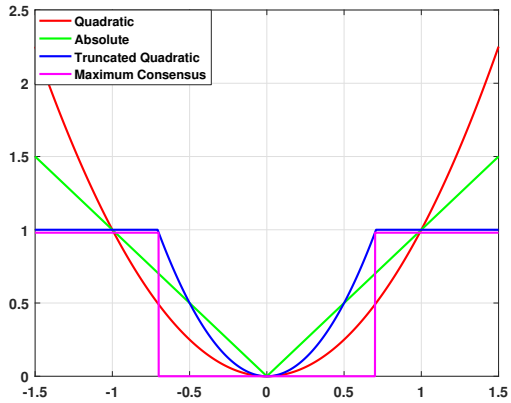


Fig. 1: The family of loss functions: quadratic (ℓ_2) loss, absolute (ℓ_1) loss, truncated quadratic (truncated ℓ_2) loss and maximum consensus loss.

there are a considerable number of robust loss functions proposed in the computer vision literature, we focus on only two types of them in this paper that exhibit one of the properties. We note that in the ideal case with noiseless data, ℓ_0 loss characterizes the most effective robust function. As a generalization of the ℓ_0 loss, the first type of interest is the group of truncated losses which can accommodate noisy data and result in an exact solution. The property of convexity is also very important because it has huge implications for the hardness of searching the optimal solution. In this paper, the second type of interest is the ℓ_1 loss, which can be seen as a very tight convex relaxation of the exact truncated loss. We visualize the representative loss functions in Fig. 1. We show that for the two types of robust loss functions efficient solvers can be developed that run in milliseconds to perform deterministic search. In the context of geometric model fitting, we accordingly propose a class of algorithms based on the introduced efficient solvers. Theoretical and experimental analyses are made for recommendations of the proposed algorithms. Qualitative and quantitative experiments are conducted on publicly available datasets and benchmarks which demonstrate the superiority of the proposed methods in comparison with the state-of-the-art competitors.

Briefly speaking, the major novelty and contribution of this paper are to provide a new perspective for geometric model fitting guided by robust loss functions (*i.e.* truncated loss and ℓ_1 loss). In this regard, optimization algorithms are also proposed which can be seen as efficient realizations of our robust loss framework. More concretely, the contributions of this paper can be summarized as follows:

- We provide an analysis for the two important types of robust loss functions and develop efficient solvers for both. Specifically, for the case of ℓ_1 loss, we relate it to the field of robust subspace recovery to propose an efficient *projected sub-gradient descent* solver. For the case of truncated losses, we propose an efficient solver based on deterministic annealing to handle the non-convexity.
- In the context of geometric model fitting, *i.e.* homography and fundamental matrix estimation, a class of algorithms are designed leveraging the efficient solvers induced by the robust loss functions. Recommendations of the best performers are made based on theoretical and experimental analyses.
- Extensive experiments are conducted on publicly available datasets and benchmarks for homography estimation, fundamental matrix estimation and the downstream task of wide-baseline stereo, which demonstrate the superiority of our method against the state-of-the-art methods.

A preliminary version of this manuscript appears in [16]. This paper is a comprehensive extension of the conference version. The idea of designing efficient algorithms based on robust loss function has been generalized, and the study of truncated losses is newly included in this paper. The study of an ℓ_1 loss-based geometric fitting method is reorganized and an additional and theoretically more grounded case for detecting homography-related correspondences is discussed. Moreover, the experimental results are significantly extended to further analyze the property of the proposed methods, and demonstrate their efficacy on more benchmarks.

The remainder of this paper is organized as follows. Section 2 describes background material and related work. In Section 3, we propose our method by presenting the investigation for truncated loss and ℓ_1 loss, respectively. Section 4 compares different formulations of the proposed method and illustrates the performance of our method in comparison with other approaches on different datasets and benchmarks. In Section 5, we summarize the paper with some concluding remarks.

2 Related Work

There is a large volume of methods in the literature proposed to address the geometric model fitting problem. Since a comprehensive review that covers all branches is exhaustive and out of the scope of this paper, in this section, we only summarize the closely related work that puts our paper into context.

Due to the practical demand of both robustness and efficiency for geometric model fitting, the random sampling techniques remain to be the most prevalent paradigm. A large number of innovations have been proposed in the past few decades to advance the primitive RANSAC [6], in terms of both efficiency and accuracy. For acceleration, many efficient sampling techniques have been proposed, taking advantage of the prior information available in feature correspondences. For example, as priors, spatial coherence is utilized in NAPSAC [17] and GroupSAC [18], and matching scores in EVSAC [19] and PROSAC [8]. Moreover, improving the model verification stage using randomized strategies has also been shown to be critical to reduce the computational cost without sacrificing the robustness, such as SPRT [20] and $T_{d,d}$ test [21].

There are also some efforts that have proven to be able to obtain more accurate estimation results. These methods include MLESAC [22] and MAPSAC [23], in which the model quality is evaluated with a maximum likelihood process. A more illuminating idea is proposed in locally optimized RANSAC (LO-RANSAC) [7], [24], where a local optimization step is introduced to polish the so-far-the-best

model. By involving more inliers for estimation, the bias induced from noises is reduced and a more accurate model can be expected. For estimation of the epipolar geometry, degeneracy in solution is a problem that cannot be ignored. This is addressed by DEGENSAC [25], which introduces an additional step to test the degeneracy and perform effective sampling. Notably, by combining the most promising improvements, USAC [26] is proposed as the state-of-the-art RANSAC variant. More recently, GC-RANSAC [27] has been proposed using a graph-cut algorithm in the local optimization step, leading to more accurate estimates of geometric models. Additionally, MAGSAC [28] and MAGSAC++ [29] are proposed to eliminate the need for a predefined inlier-outlier threshold which is critical but hard to tune in practice.

From a different perspective, there also exists a large group of deterministic search algorithms with an optimization-based formulation. Usually, the primary objective is known as *consensus maximization*, which stems from the model quality evaluation strategy of RANSAC, *i.e.* counting the number of correspondences with the residuals below a given threshold. In this regard, RANSAC can be seen as a stochastic solver with no guarantee of the quality of solution. A variety of methods attempt to develop algorithms to search the solution with global optimality guarantee, using techniques such as branch-and-bound [9], [10], [30], tree search [11], [31], or enumeration [32], [33]. Recent years have witnessed a surge of algorithms that attempt to optimize the more robust maximum consensus objective, albeit approximately or asymptotically. In [34], the objective is relaxed to a smooth surrogate function to avoid local solutions. In [14], two deterministic approximate approaches are proposed, and the objective is asymptotically approached by an exact penalty method and an Alternating Direction Method of Multiplier (ADMM) technique, respectively. Additionally, in [13], a biconvex programming technique is introduced to forcibly increase the consensus. These methods have exhibited promising improvements over randomized methods, giving consistently higher consensus. However, without exception in these methods, the optimization procedure is decomposed into sub-problems such as Linear Program (LP), Second Order Cone Program (SOCP) or Quadratic Program (QP), where convex solvers are required. This characteristic incurs great computational cost and reduces their practicality.

Although the geometric models can characterize the problem in a simple and explicit way, and thus are crucial to guide the design of algorithms, there has been a considerable number of methods using different formulations without leveraging the parametric models. Instead, some other statistically meaningful priors in image matching are utilized. This includes the VFC method [35], which prunes the outliers by recovering a smooth function in a Reproducing Kernel Hilbert Space (RKHS). Similarly, the CODE algorithm [36] uses non-linear regression techniques to impose the motion smoothness constraint for discovering consistent matches. Recent advances have suggested more efficient algorithms to prune the erroneous matches with a smoothness constraint, such as LPM [37] and GMS [38], reporting promising results. In addition, deep learning techniques have also been examined for geometric model fitting. The method learning to find good correspondences (LFGC) [39] has been proposed as a first attempt. It trains a multi-layer perceptron-based deep network to label the correspondences. The method has also encouraged several following-up works that have been shown to be more effective in the deep learning framework [40], [41], [42]. However, these methods address the geometric model fitting problem using priors that are either only statistically meaningful or learned as a black-box from data, which restricts their generalization.

Notably, the ℓ_1 loss-based fitting problem is closely related to the field of *Robust Subspace Recovery*, as we will explain in the next section. There is a rich literature in this field, including some non-convex heuristic solutions [43], [44], [45] and theoretically justified convex relaxations [46], [47]. The interested readers are referred to the comprehensive survey in [48]. In addition, a recent paper [49] has extended the method of [43] to the geometric fitting problem in computer vision, yet only applicable to homography estimation.

3 METHODOLOGY

This paper is designated to study the problem of geometric model fitting. In particular, suppose we are given a set of tentative 2D image correspondences $S = \{(\mathbf{x}_i, \mathbf{x}'_i)\}_{i=1}^N$ with a number of outliers, where $\mathbf{x}_i = (x_i, y_i, 1)^T$ and $\mathbf{x}'_i = (x'_i, y'_i, 1)^T$ are column vectors denoting the homogeneous coordinates of feature points from two images, our primary aim is to recover the underlying geometric structure, such as the fundamental matrix $\mathbf{F} \in \mathbb{R}^{3\times3}$ or homography $\mathbf{H} \in \mathbb{R}^{3\times3}$ that is essential for many 3D vision applications. The properties of the geometric models $\mathbf{F}$ and $\mathbf{H}$ are well-known and their correlations *w.r.t.* image correspondences are also general knowledge [5]. In the outlier-free case, the method for estimating the models has been well-studied.

Theoretically, the distribution of data determines the optimal algorithm for estimation. In our context, it suggests that the primary objective to optimize should be expressed as a geometric quantity (*i.e.* geometric error), which is the canonical view for geometric model fitting. However, this generally induces highly complex non-linear objectives which are computationally very expensive to find the optimal solution. Fortunately, it has been shown that the linearized error (*i.e.* algebraic error) is very effective if the data are properly normalized, leading to much simpler algorithms. In this paper, since our aim is to develop a method that works in practical settings, we concentrate on the study of algebraic errors.

The efficient solution using linearized error for geometric estimation in the outlier-free case is known as the Direct Linear Transformation (DLT) method [5]. Before elaborating on our method, we provide an overview of the general structure of the geometric models, as well as the basic DLT method which has a strong connection to ours. As will be seen, our method can be seen as an extension of DLT in several fundamental aspects.

The **fundamental matrix $\mathbf{F}$** governs the most general epipolar constraint in two camera views. This constraint can be expressed as:

$$\mathbf{x}_i'^T \mathbf{F} \mathbf{x}_i = 0, \tag{1}$$

in which $\mathbf{F}$ is of rank 2 and has 7 degrees of freedom up to scale. Given sufficient correspondences (at least 8 for DLT), Eq. (1) can be used to compute the unknown matrix $\mathbf{F}$. Denote by $\mathbf{f}$ the 9-vector made up of the entries of $\mathbf{F}$ in row-major order, Eq. (1) can be expressed as a vector inner product:

$$\mathbf{a}_i^T \mathbf{f} = 0, \tag{2}$$

where

$$\mathbf{a}_i = (x'_i x_i, x'_i y_i, x'_i, y'_i x_i, y'_i y_i, y'_i, x_i, y_i, 1)^T, \tag{3}$$

represents an embedding of the correspondence data. Thus, given n outlier-free correspondences ($n \geq 8$), DLT obtains the solution by solving the following overdetermined linear system: $\mathbf{Mf} = \mathbf{0}$, where $\mathbf{M} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]^T$. Throughout the paper, we abuse the notation $\mathbf{M}$ to represent data matrix formed from the correspondence data. The solution is given as the right singular vector corresponding to the smallest singular value of $\mathbf{M}$.

The **homography transformation** applies when the feature points are lying close to a plane or the camera motion is a pure rotation. The transformation can be expressed in terms of the vector cross product:

$$\mathbf{x}_i'^T \times \mathbf{H}\mathbf{x}_i = \mathbf{0}, \tag{4}$$

in which $\mathbf{H}$ is non-singular with 8 degrees of freedom. In this case, each correspondence gives rise to two independent equations. Denote by $\mathbf{h}$ the 9-vector made up of the entries of $\mathbf{H}$ in the row-major order, and we use the notation $\mathbf{b}_i$ to represent the embedding, the following equation holds:

$$\mathbf{b}_i^T \mathbf{h} = \mathbf{0}, \tag{5}$$

where $\mathbf{b}_i$ is given by

$$\mathbf{b}_i^T = \begin{bmatrix} \mathbf{0}^T & -\mathbf{x}_i^T & y'_i \mathbf{x}_i^T \\ \mathbf{x}_i^T & \mathbf{0}^T & -x'_i \mathbf{x}_i^T \end{bmatrix}. \tag{6}$$

Given n correspondences ($n \geq 4$), Eq. (5) also enables a linear system $\mathbf{Mh} = \mathbf{0}$ to determine homography with $\mathbf{M} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n]^T$. Analogously, the solution is given as the right singular vector corresponding to the smallest singular value of $\mathbf{M}$.

In summary, the DLT method for outlier-free geometric model fitting can be characterized as the following optimization problem:

$$\min_{\|\mathbf{z}\|=1} \|\mathbf{Mz}\|_2^2, \tag{7}$$

where the data matrix $\mathbf{M}$ is composed of the specific embeddings of correspondences, and $\mathbf{z}$ denotes the vector of parameters. The analytic solution is exactly the right singular vector corresponding to the smallest singular value of $\mathbf{M}$.

The importance of Problem (7) is that it can to some extent explain the limitation of the DLT method and also encourages a more general view for geometric model fitting. Intuitively, the reason that DLT can only be applied in the outlier-free case is that the objective in (7) is an ℓ_2-based error, which is not robust. Thus a straightforward innovation is to generalize the problem to objectives based on robust loss functions, which will be studied shortly in what follows.

3.1 Exact Solution with Truncated Loss

3.1.1 Robust Formulation with Truncated Loss

In (almost) all practical computer vision tasks, the acquired data are imperfect and contaminated by noise, outliers or both. Considering both the challenges posed by noise and outliers, exact estimation can be approached by using truncated losses, since they can diminish the influence of outliers and tolerate noise in a given level. Notice that there is a strong link between the principle to maximize the consensus size and to optimize over truncated losses. Typically, the principle of *consensus maximization* can be seen as minimizing the maximum consensus loss. We can see that from a statistical view, the maximum consensus loss and different truncated losses are merely distinct in the assumption of the noise for inliers. For instance, truncated ℓ_2 loss is arguably the optimal choice given that the noise of inliers is subject to a Gaussian distribution, as suggested in MLESAC [22]. Similarly, maximum consensus loss assumes that the noise is subject to a uniform distribution. We note that although these losses are theoretically distinct, their performances on real-world data are only marginally different given the noise level [24]. Thus the main concern here is the numerical solvers that can address the non-convexity of the problem. As previously mentioned, the maximum consensus loss has been considered by consensus maximization methods, which are generally computationally expensive. In this paper, we propose a novel efficient solver for robust geometric model fitting with truncated losses, which draws inspiration from deterministic annealing to optimize the objective in a maximum entropy principle to cope with the non-convexity.

Although other interpretations are available, in this paper, we start developing our algorithm by introducing consensus maximization. In essence, the consensus maximization problem is to find the model that is consistent with the most inliers in the data. The inliers are defined up to a given inlier-outlier threshold over the residual, which is in turn determined by the model.

Given N correspondences, the consensus maximization problem for geometric model fitting can be defined as:

$$\max_{\|\mathbf{z}\|=1} \sum_i \mathbb{I}(r_i(\mathbf{z}) \leq \epsilon), \tag{8}$$

where $\mathbb{I}(\cdot)$ is the indicator function that returns 1 if its input condition is true and 0 otherwise, ϵ is the given inlier-outlier threshold. As aforementioned, $\mathbf{z}$ denotes the vector of model parameters, $r_i(\mathbf{z})$ denotes the non-negative residual of the i-th correspondence $\mathbf{x}_i$ *w.r.t.* model $\mathbf{z}$. The constraint $\|\mathbf{z}\| = 1$ is to avoid trivial solutions such as $\mathbf{z} = \mathbf{0}$. This can be equivalently expressed as optimizing with the maximum consensus loss:

$$\begin{aligned} \ell_c(r_i(\mathbf{z})) &= \begin{cases} 0, & \text{if } r_i(\mathbf{z}) \leq \epsilon, \\ 1, & \text{otherwise,} \end{cases} \\ &= (1 - p_i), \end{aligned} \tag{9}$$

where $p_i \in \{0, 1\}$ denotes the indicator variable of each correspondence $(\mathbf{x}_i, \mathbf{x}'_i)$, *i.e.* $p_i = 1$ indicates that $(\mathbf{x}_i, \mathbf{x}'_i)$ is an inlier satisfying $r_i(\mathbf{z}) \leq \epsilon$, and $p_i = 0$ otherwise. We consider the minimization form of consensus maximization,

which is straightforwardly related to the loss function:

$$\min_{\|\mathbf{z}\|=1,\mathbf{p}} \sum_i (1-p_i), \quad s.t.\ p_i r_i(\mathbf{z}) \leq \epsilon,\ \ p_i \in \{0,1\}, \quad (10)$$

where $\mathbf{p} = [p_1, p_2, \ldots, p_N]^T$. The equivalence between (8) and (10) can be easily established, since if we have $p_i = 1$, the necessary condition $r_i(\mathbf{z}) \leq \epsilon$ is forcibly satisfied by the constraint $p_i r_i(\mathbf{z}) \leq \epsilon$.

In this paper, instead of the maximum consensus loss in (9), we consider the group of truncated losses in the form:

$$\begin{aligned} \ell_t(r_i(\mathbf{z})) &= \begin{cases} f(r_i(\mathbf{z})), & \text{if } r_i(\mathbf{z}) \leq \epsilon, \\ 1, & \text{otherwise}, \end{cases} \\ &= p_i \cdot f(r_i(\mathbf{z})) + (1-p_i), \end{aligned} \quad (11)$$

where $f(\cdot)$ represents a commonly used loss function, *e.g.* ℓ_1, ℓ_2, or hinge loss, *etc*. We note that the truncated loss ℓ_t should be continuous, thus a basic property for $f(\cdot)$ is that $f(\epsilon) = 1$, which can be approached by including a scaling factor in the loss function. For example, if we take ℓ_2 loss, the exact form of $f(\cdot)$ will be $f(x) = \frac{1}{\epsilon^2}x^2$, with x being the variable and $\frac{1}{\epsilon^2}$ as the scaling factor.

The decomposition (11) leads to a relaxed form of Problem (10), given as:

$$\begin{aligned} \min_{\|\mathbf{z}\|=1,\mathbf{p}} \sum_i \Big(p_i \cdot f(r_i(\mathbf{z})) + (1-p_i)\Big), \\ s.t.\ \ p_i r_i(\mathbf{z}) \leq \epsilon,\ \ p_i \in \{0,1\}. \end{aligned} \quad (12)$$

At a first glance, Problem (12) is still very complex and hard to optimize due to the high-order and discrete constraints. However, we can alternatively solve the following problem to find a stationary point for (12):

$$\min_{\|\mathbf{z}\|=1,\mathbf{p}} \sum_i \Big(p_i \cdot f(r_i(\mathbf{z})) + (1-p_i)\Big), \quad s.t.\ \ p_i \in [0,1]. \quad (13)$$

Note that ϵ implicitly appears in (13) since we enforce $f(\epsilon) = 1$. This is supported by the following observation:

Proposition 1: *All local minima of Problem* (13)*, if we let* $p_i = 1$ *in case* $r_i(\mathbf{z}) = \epsilon$*, satisfy the constraints in* (12)*.*

The proof of **Proposition 1** is simple. First, it is worth mentioning that to provide a tight approximation to the maximum consensus loss, we must have $f(\epsilon) = 1$. We then can derive that $f(t) < 1$ indicates $t < \epsilon$ for $t \in [0, \epsilon]$. If $\mathbf{z}$ is fixed, (13) reduces to a simple linear assignment problem. Obviously, if $r_i(\mathbf{z}) \neq \epsilon$, we must have $p_i = 1$ or 0. We have $f(r_i(\mathbf{z})) < 1$ for $p_i = 1$ and $f(r_i(\mathbf{z})) > 1$ for $p_i = 0$, which equals to the constraint $p_i r_i(\mathbf{z}) \leq \epsilon$. Practically, (13) is a continuous optimization problem with linear constraints, which admits efficient solutions as will be shown.

3.1.2 Efficient Solver for Truncated Loss

Analogously to the maximum consensus loss, truncated losses are also highly non-convex. This induces a great difficulty to find a global solution for Problem (13). To this end, we propose a deterministic annealing method based on a maximum entropy principle [50].

In brief, deterministic annealing introduces fuzzy assignment in the optimization process, which is achieved by adding an entropy term for regularization. In our case, the problem is translated into the form:

$$\begin{aligned} \min_{\|\mathbf{z}\|=1,\mathbf{p}} \sum_i \Big(p_i \cdot f(r_i(\mathbf{z})) + (1-p_i)\Big) \\ +\alpha \sum_i \Big(p_i \log p_i + (1-p_i)\log(1-p_i)\Big), \\ s.t.\ \ p_i \in [0,1], \end{aligned} \quad (14)$$

where α is the temperature parameter. The deterministic method first uses a large α in the initial stage, then decrease it to 0 in the optimization process. When α approaches 0, (14) degenerates to the original form.

We can leverage an alternating minimization algorithm to solve (14). Basically, the algorithm alternates between solving two sub-problems:

Q1: Finding $\mathbf{p}$ given $\mathbf{z}$,

$$\begin{aligned} \min_{\mathbf{p}} \sum_i \Big(p_i \cdot f(r_i(\mathbf{z})) + (1-p_i)\Big) \\ +\alpha \sum_i \Big(p_i \log p_i + (1-p_i)\log(1-p_i)\Big), \\ s.t.\ \ p_i \in [0,1], \end{aligned} \quad (15)$$

Q2: Finding $\mathbf{z}$ given $\mathbf{p}$,

$$\min_{\|\mathbf{z}\|=1} \sum_i p_i \cdot f(r_i(\mathbf{z})). \quad (16)$$

The procedure iterates until a local optimum is approached. The two sub-problems are much easier to address, resulting in efficient solutions.

It is straightforward to see that without the regularization term, Problem (15) will be a simple linear assignment problem. In fact, the problem has a closed-form solution as:

$$p_i = \begin{cases} 1, & \text{if } f(r_i(\mathbf{z})) \leq 1, \\ 0, & \text{otherwise}. \end{cases} \quad (17)$$

Clearly, p_i will be always discrete taking a value in $\{0,1\}$. In this case, the search is confined to discrete domain, and may very likely get trapped in a poor local optimum. In the deterministic annealing framework, the situation is quite different as we will see. Taking the derivative of the objective in (15) yields the following stationary condition:

$$p_i = \frac{\exp(-\frac{f(r_i(\mathbf{z}))}{\alpha})}{\exp(-\frac{f(r_i(\mathbf{z}))}{\alpha}) + \exp(-\frac{1}{\alpha})}. \quad (18)$$

It can be seen that $p_i \in [0,1]$ is naturally satisfied. At higher temperatures, the entropy term forces p_i to be more fuzzy. The minima obtained at each temperature are used as initial conditions for the next stage as the temperature is lowered. Clearly, as the temperature decreases to 0, the problems reduce to hard assignment as in (17). The update of p_i can also be explained as a *softmax* operation. From an optimization point of view, the fuzziness of p_i raised by softmax makes the resulting energy function better behaved because the objective is able to be improved gradually and continuously during the optimization without jumping around in the discrete space.

Remark 1: *In our loss decomposition framework* (11)*, the loss*

Algorithm 1: Geometric Model Fitting with Truncated ℓ_2 Loss

Input: The correspondence set S, parameters α_0, γ, ϵ
Output: The vector of parameters $\mathbf{z}$
1: Initialize $\mathbf{z}$, $\alpha = \alpha_0$.
2: **while** objective in Problem (13) not converge **do**
3: Update $\mathbf{p}$ using the Eq. (18).
4: Update $\mathbf{z}$ using the closed-form solution of (20) or (21).
5: Annealing: $\alpha = \gamma\alpha$.
6: **end while**
7: Return optimal model $\mathbf{z}$.

function $f(\cdot)$ can take many forms such as ℓ_1, ℓ_2, hinge loss, etc. *To achieve the best approximation of the maximum consensus loss,* i.e. *ℓ_c, the ideal case is to adopt hinge loss. By tuning the threshold h of hinge loss such that $h \to \epsilon$, the robust truncated loss ℓ_t is able to approximate ℓ_c by arbitrary accuracy. For Problem* (16)*, by taking hinge loss with parameter h, it becomes a generalized hinge loss regression problem given as:*

$$\min_{\|\mathbf{z}\|=1} \sum_i p_i \cdot \xi_i, \quad s.t.\ r_i(\mathbf{z}) \leq h + \xi_i,\ \ \xi_i \geq 0, \tag{19}$$

where ξ_i is the slack variable. Nevertheless, this problem fundamentally excludes an efficient solution. In this paper, we concentrate on truncated ℓ_2 loss due to the efficient global solution permitted.

To solve Problem (16), it requires us to specify the exact form of $r_i(\mathbf{z})$. Essentially, by adopting the algebraic error, the analytical problem can be readily derived by basic linear algebra, as we will show in the following.

For the estimation of the homography model, we have $r_i(\mathbf{h}) = \|\mathbf{b}_i^T\mathbf{h}\|_2$, Problem (16) can be expressed in a concise form as:

$$\min_{\|\mathbf{h}\|=1} \sum_i p_i \|\mathbf{b}_i^T\mathbf{h}\|_2^2 = \mathbf{h}^T\mathbf{M}^T\hat{\mathbf{P}}\mathbf{M}\mathbf{h}, \tag{20}$$

where $\mathbf{M} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n]^T$, $\mathbf{P} = diag(\mathbf{p})$ and $\hat{\mathbf{P}} = \mathbf{P} \otimes \mathbf{I}_{2\times 2}$, with $\otimes$ denoting the Kronecker product operator. The closed-form solution of (20) is the eigenvector of matrix $\mathbf{M}^T\hat{\mathbf{P}}\mathbf{M}$ with least eigenvalue, or equivalently the right singular vector corresponding to the smallest singular value of $\hat{\mathbf{P}}^{\frac{1}{2}}\mathbf{M}$.

For the estimation of the fundamental matrix, we have $r_i(\mathbf{f}) = |\mathbf{a}_i^T\mathbf{f}|$, Problem (16) can be expressed in a concise form as:

$$\min_{\|\mathbf{f}\|=1} \sum_i p_i |\mathbf{a}_i^T\mathbf{f}|^2 = \mathbf{f}^T\mathbf{M}^T\mathbf{P}\mathbf{M}\mathbf{f}, \tag{21}$$

where $\mathbf{M} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]^T$. Analogously, the closed-form solution of (21) is the eigenvector of matrix $\mathbf{M}^T\mathbf{P}\mathbf{M}$ with the least eigenvalue, or equivalently the right singular vector corresponding to the smallest singular value of $\mathbf{P}^{\frac{1}{2}}\mathbf{M}$.

This concludes our algorithm for the efficient solution using robust truncated losses, since Problem (15) and Problem (16) are solved with closed-form solution. We summarize the algorithm in Alg. 1.

3.2 Relaxed Solution with Convex ℓ_1 Loss

As previously mentioned, an exact solution can be approached by using a truncated loss, which suffers from the non-convexity of the loss function. In contrast, some other loss functions such as ℓ_1 loss and *Huber* loss are convex, which can be seen as the relaxed form of the exact losses. Note that the convexity of a loss function does not necessarily indicate that the optimization problem is convex, but an easier optimization problem can be expected.

In this subsection, we focus on the study of ℓ_1 loss for geometric model fitting. As will be shown, the induced optimization problem can be related to *robust subspace recovery*, and recent advances have shown that an efficient *projected gradient-descent* based solver can suffice the requirements for geometric model fitting.

Recall from (7), and replace ℓ_2 loss with ℓ_1 loss, we have a simple form as follows for geometric model fitting:

$$\min_{\|\mathbf{z}\|=1} \ \|\mathbf{M}\mathbf{z}\|_1. \tag{22}$$

Mathematically, (22) can be seen as a hyperplane fitting problem. In fact, the exact form of (22) has been recently investigated in the literature of robust subspace recovery [43], [51], where hyperplane fitting is a special case when the intrinsic dimension of data $d = D - 1$, with D representing the ambient dimension of the data. The robust property has been theoretically demonstrated, which roughly states that under some assumptions on the distributions of outliers, the estimation task with (22) can even tolerate $O(m^2)$ outliers, where m denotes the inlier number.

Note that (22) is non-convex (since the feasible region is a sphere) and non-smooth (due to the ℓ_1-based objective), therefore the solution is non-trivial and needs additional care. Fortunately, several efforts on the numerical solver for (22) have been proposed. In [43], (22) is relaxed to a sequence of *linear programs*, which guarantees finite convergence to the global optima. However, this approach is computationally expensive. Alternatively, [43] provides an *iteratively reweighted least squares*-based method, which is more efficient but comes with no theoretical guarantees. A *projected sub-gradient descent*-based algorithm is proposed in [51]. The algorithm is even more efficient involving only matrix-vector multiplications. Since the demand for low computational time usually dominates the need of optimality guarantees for geometric estimation, we adopt the projected sub-gradient descent-based algorithm. The theoretical performance of this algorithm under outliers as well as noise is studied in [52].

In the context of this paper, the task of geometric model fitting can benefit from (22) in a straightforward way. All we need to do is finding a proper data matrix $\mathbf{M}$ from the given correspondences (as in the DLT method), and essentially (22) will be searching the parameters by minimizing the ℓ_1-based algebraic error. We outline the proposed geometric estimation method with (22) in Alg. 2.

We note that since (22) is still non-convex, it also suffers from the issue of weak local optima. We argue that from a different point of view, this risk can be further reduced. In subspace learning theory, it is well-known that the relative dimension, *i.e.* d/D, the quotient of intrinsic dimension of data d and the dimension of ambient space

Algorithm 2: Geometric Model Fitting with ℓ_1-based Hyperplane Fitting

Input: The correspondence set $\mathcal{S}$
Output: The orthogonal vector $\mathbf{z}$ of the sought hyperplane
1: Mapping correspondences into embeddings $\mathbf{a}_i$ or $\mathbf{b}_i$ to form the data matrix $\mathbf{M}$.
2: Initialize $\mathbf{z}$ as the right singular vector corresponding to the smallest singular value of $\mathbf{M}$.
3: **while** not converge **do**
4: Compute sub-gradient: $\mathbf{g} = \mathbf{M}^T sign(\mathbf{M}\mathbf{z})$.
5: Update step size μ according to a certain rule [51].
6: Sub-gradient descent: $\mathbf{z} \leftarrow \mathbf{z} - \mu\mathbf{g}$.
7: Sphere projection: $\mathbf{z} \leftarrow \mathbf{z}/\|\mathbf{z}\|_2$.
8: **end while**
9: Return $\mathbf{z}$.

D, plays an important role in the difficulty of the learning task. Generally speaking, the subspace learning problem is significantly easier when the relative dimension is small. In geometric model fitting, we show that this fact can be fruitfully exploited by considering a simpler embedding of correspondences.

In this view, we will show that the embedding $\mathbf{b}_i$ in (6) used by DLT for homography estimation is unfavored and another formulation is inherently easier. The new formulation is based on the following observation:

Proposition 2: *Given $n \geq 4$ correspondences with no noise and outliers and conforming to a projective transformation $\mathbf{H}$, the 9-dimensional embeddings $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ live in a linear subspace with dimension no more than 6.*

Proof. It can be derived from Eq. (4) using linear algebra that

$$\mathbf{H}'\mathbf{a}_i = 0 \tag{23}$$

holds for $\forall i = 1, 2, \ldots, n$, where

$$\mathbf{H}' = \begin{bmatrix} \mathbf{0}^T & \mathbf{h}_3^T & -\mathbf{h}_2^T \\ -\mathbf{h}_3^T & \mathbf{0}^T & \mathbf{h}_1^T \\ \mathbf{h}_2^T & -\mathbf{h}_1^T & \mathbf{0}^T \end{bmatrix} \tag{24}$$

for $\mathbf{H} = [\mathbf{h}_1^T; \mathbf{h}_2^T; \mathbf{h}_3^T]$. It can be seen from (23) and (24) that $\mathbf{a}_i$ lives in a linear subspace. Since $\mathbf{H}'$ is clearly of rank 3, the dimension of the linear subspace is no more than 6. □

As we will show later, **Proposition 2** leads to a subspace recovery problem for homography estimation. Following this idea, the simplest embedding of correspondences with only first-order terms

$$\mathbf{d}_i = [x_i, y_i, x_i', y_i', 1]^T \tag{25}$$

is also of great interest. Intuitively, this embedding is related to the affine model which is a linear transformation. Assume that the correspondences are related by the affine model:

$$\mathbf{x}_i' = \mathbf{A}\mathbf{x}_i, \tag{26}$$

where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \tag{27}$$

represents the affine matrix. The structure of $\mathbf{d}_i$ is revealed by the following proposition:

Proposition 3: *Given $n \geq 3$ correspondences with no noise and outliers and conforming to an affine transformation $\mathbf{A}$, the 5-dimensional embeddings $\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n$ live in a linear subspace with dimension no more than 3.*

Proof. It can be derived from Eq. (26) that

$$\mathbf{A}'\mathbf{d}_i = 0 \tag{28}$$

holds for $\forall i = 1, 2, \ldots, n$, where

$$\mathbf{A}' = \begin{bmatrix} a_{11} & a_{12} & -1 & 0 & a_{13} \\ a_{21} & a_{22} & 0 & -1 & a_{23} \end{bmatrix}. \tag{29}$$

It can be seen from (28) and (29) that $\mathbf{d}_i$ lives in a linear subspace. Since $\mathbf{A}'$ is clear of rank 2, the dimension of the linear subspace is no more than 3. □

Remark 2: *Analogous to the homography estimation case of DLT, a straightforward solution to leverage this structure is to transform it into a hyperplane fitting problem, with the following embedding:*

$$\mathbf{c}_i^T = \begin{bmatrix} \mathbf{x}_i^T & \mathbf{0}^T & -x_i' \\ \mathbf{0}^T & \mathbf{x}_i^T & -y_i' \end{bmatrix}. \tag{30}$$

The problem can be then readily solved using (22) *given $n \geq 3$ correspondences, with $\mathbf{M} = [\mathbf{c}_1^T, \mathbf{c}_2^T, \ldots, \mathbf{c}_n^T]^T$ and $\mathbf{z} = [a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}, 1]^T$ encoding the affine parameters. The benefits of leveraging $\mathbf{d}_i$ will be explained next.*

Both **Proposition 2** and **Proposition 3** suggest solving the following subspace recovery problem instead of the hyperplane fitting alternative:

$$\min_{\mathbf{v} \in \mathbb{R}^{D \times k}} \ \sum_i \|\mathbf{e}_i^T\mathbf{v}\|_1 = \|\mathbf{M}\mathbf{v}\|_{1,1}, \quad s.t. \ \mathbf{v}^T\mathbf{v} = \mathbf{I}, \tag{31}$$

where $\mathbf{M} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n]^T$, $\mathbf{e}_i$ represents the embedding of data, and $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$ represents the matrix of orthogonal unit vectors, $\mathbf{I}$ represents the identity matrix, and $\|\cdot\|_{1,1}$ represents the sum of the ℓ_1 norms of the rows of the input matrix. For homography and affine estimation, the ambient dimension for (31) is 9 and 5 respectively, and the dimension of the subspace is 6 ($k = 3$) and 3 ($k = 2$), respectively. The relative dimension is then $6/9$ and $3/5$ for (31), which is much smaller than $8/9$ and $6/7$ indicated by the hyperplane fitting case. This renders the problem a much easier task for learning.

The rationale behind (31) is to find k bases of the orthogonal complement of the linear subspace spanned by the embeddings of inliers. This can be solved by standard robust subspace recovery methods, *e.g.* [46], as discussed in the comprehensive survey [48]. In this paper, we adopt a more efficient strategy to iteratively search the k bases. In the first iteration, a hyperplane fitting algorithm is conducted to find the first basis. In the second to k-th iteration, the procedure is similar to hyperplane fitting but with an additional projection step to find the basis. The additional projection step guarantees that the sought basis is orthogonal to the previous ones. Specifically, if we have obtained the bases $\mathbf{v}_p \in \mathbb{R}^{D \times i}$, the projector of its orthogonal complement should be $\mathbf{I} - \mathbf{v}_p\mathbf{v}_p^T$, then the sought basis $\mathbf{v}_{i+1}$ should be projected onto it as $\mathbf{v}_{i+1} = (\mathbf{I} - \mathbf{v}_p\mathbf{v}_p^T)\mathbf{v}_{i+1} = \mathbf{v}_{i+1} - \mathbf{v}_p\mathbf{v}_p^T\mathbf{v}_{i+1}$. The algorithm to solve (31) is outlined in Alg. 3.

Algorithm 3: Geometric Model Fitting with ℓ_1-based Subspace Recovery

Input: The correspondence set $\mathcal{S}$
Output: The bases of the orthogonal complement $\mathbf{v}$

Mapping the correspondences to form the data matrix $\mathbf{M}$.
Initialize $\mathbf{v}_0 = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$ as the right singular vectors of the two smallest singular values of $\mathbf{M}$.
for $i = 1$ to k **do**
  while not converge **do**
    Compute sub-gradient: $\mathbf{g}_i = \mathbf{M}^T sign(\mathbf{M}\mathbf{v}_i)$.
    Update the step size ν according to a certain rule [51].
    Sub-gradient descent: $\mathbf{v}_i \leftarrow \mathbf{v}_i - \nu_j \mathbf{g}_i$.
    if $i > 1$ **then**
      Orthogonal projection: $\mathbf{v}_i \leftarrow \mathbf{v}_i - \mathbf{v}_p \mathbf{v}_p^T \mathbf{v}_i$.
    end if
    Sphere projection: $\mathbf{v}_i \leftarrow \mathbf{v}_i / \|\mathbf{v}_i\|_2$.
  end while
  Store current bases: $\mathbf{v}_p = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i]$.
end for
return $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$.

Algorithm 4: Homography Estimation Based on ℓ_1 Loss

Input: The correspondence set $\mathcal{S}$
Output: The estimated model $\mathbf{H}$

Apply Alg. 2 or Alg. 3 on $\mathcal{S}$ to find the orthogonal base(s) of the subspace spanned by inliers.
Compute the residuals and apply thresholding to find the potential inlier set $\mathcal{I}$.
Post-processing on $\mathcal{I}$ to determine final estimation result $\mathbf{H}$.
return $\mathbf{H}$.

Remark 3: *Different from the exact truncated losses, a post-processing stage is necessary for our ℓ_1 loss based estimation. This is because (i) the ℓ_1 loss is inexact in nature; (ii) the subspace recovery problem* (31) *does not explicitly output the model parameters. Thus,* (22) *or* (31) *functions as detecting the inliers or rejecting the outliers, which requires a post-processing stage to obtain the model parameters.*

To conclude, for ℓ_1-based geometric model fitting, the hyperplane fitting problem (22) can be used with embedding $\mathbf{a}_i$ for fundamental matrix estimation, or with $\mathbf{b}_i$ for homography estimation. The subspace recovery problem (31) can be used with $\mathbf{a}_i$ to identify homography-related correspondences, or with $\mathbf{d}_i$ to identify affine-related correspondences. To leverage the strength of (31), we also consider the possibility to iteratively search two groups of homography-related or affine-related correspondences to avoid degeneracy for fundamental matrix estimation. The algorithms for homography and fundamental matrix estimation based on ℓ_1 loss are summarized in Alg. 4 and Alg. 5, respectively.

Algorithm 5: Fundamental Matrix Estimation Based on ℓ_1 Loss

Input: The correspondence set $\mathcal{S}$
Output: The estimated model $\mathbf{F}$

Apply Alg. 2 or Alg. 3 on $\mathcal{S}$ to find the orthogonal base(s) of the subspace spanned by inliers.
Compute the residuals and apply thresholding to find the potential inlier set $\mathcal{I}_1$.
Exclude $\mathcal{I}_1$ from $\mathcal{S}$ to form $\mathcal{S}'$.
Apply Alg. 2 or Alg. 3 on $\mathcal{S}'$ to the orthogonal base(s) of the subspace spanned by inliers.
Compute the residuals and apply thresholding to find the potential inlier set $\mathcal{I}_2$.
Post-processing on $\mathcal{I}_1 \bigcup \mathcal{I}_2$ to determine final estimation result $\mathbf{F}$.
return $\mathbf{F}$.

3.3 Implementation Details

To improve numerical stability, the correspondence data are mapped into embeddings and then normalized to unit norm before processed by our algorithm.

Post-processing for relaxed solution: The proposed method using ℓ_1 loss does not explicitly give the parameters of the sought geometric model, or the returned model can be too coarse due to the strategy of approximation and relaxation. Thus a post-processing stage is required. Since the geometric model can be recovered by a small number of samples of inliers, we adopt a random sampling scheme. Specifically, we run a fixed number of random samples (500 in our experiment) to generate hypotheses, and return the model that accommodates the most correspondences. A local optimization step is performed when a *so-far-the-best* model is sought.

Parameter Settings: For Alg. 1 with truncated ℓ_2 loss, in the annealing process, the initial temperature α_0 is set adaptively to the variance of the residuals in the first iteration, and the decay factor γ is set to 0.9. For Alg. 4, the threshold to determine the potential inliers is empirically set to 0.15. For Alg. 5, the threshold to determine the potential inliers is empirically set to 0.25 in the first iteration, and 0.15 in the second iteration. Note that more iterations can be used to enrich the detected correspondences, however, in practice we observe that two iterations are enough to avoid degeneracy and more iterations are not necessary.

4 Experimental Analyses and Results

In this section, we provide experimental investigations of the proposed method based on synthetic data and real-world datasets, which involve three following aspects. First, we conduct linear fitting experiments to study the theoretical properties of exact solution and relaxed solution. Second, as we have introduced several different formulations, their properties are studied and the best practices determined. Third, to verify the advantage of our method, experimental results are given in comparison with the state-of-the-art methods on publicly available datasets and benchmarks. We name our method with truncated ℓ_2 loss as Efficient Exact Search (EES) and method based on ℓ_1 loss as Efficient Approximate Search (EAS) for geometric model fitting.

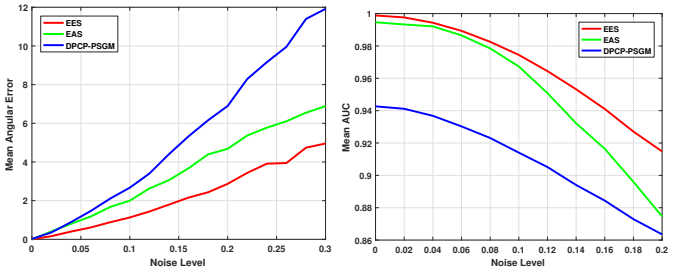


Fig. 2: Evaluation of the performances of EES, EAS, and DPCP-PSGM based on synthetic data (left) and real 3D point cloud data (right) *w.r.t.* noise level.

4.1 Theoretical Properties of Exact Solution and Relaxed Solution

We first use linear fitting experiments to investigate the properties of the exact solution and the relaxed solution using the proposed methods. Note that different from fundamental matrix and homography transformation estimation, linear fitting provides a more ideal environment to investigate the properties of solutions since we do not have to linearize the problem.

First, we use synthetic data for evaluation. We follow the settings of [51] to conduct the experiments. To generate the synthetic data, we fix the data dimension $D = 8$ and generate the inliers from a hyperplane (*i.e.* intrinsic dimension $d = 7$). This leads to a hyperplane fitting problem and can be resolved by accurately estimating the unit orthogonal vector of the unknown hyperplane. The outliers are randomly generated, and both inliers and outliers are normalized to have unit norm. We use 500 inliers and 2000 outliers to create a challenging scenario with the outlier rate of 80%. We then add Gaussian noise to the data to test the performance of EES, EAS, and the ℓ_1 loss-based method DPCP-PSGM proposed in [51]. Note that for linear fitting, the implementation of our EES and EAS is essentially the same as the case for fundamental matrix estimation, except that the embedding is the data point itself. The inlier-outlier threshold is empirically set to *max(0.05, 0.8*noise level)*. We generate 100 instances of synthetic data and report the average performance for each method. The performance is evaluated in terms of angular error (in the range of 0 to 90 degrees) between estimation of the unit orthogonal vector of the hyperplane and the ground-truth vector. The results are shown in the left plot of Fig. 2. Clearly, our EES outperforms EAS because the random sampling process in EAS is not robust to noise. Both EES and EAS outperform DPCP-PSGM, this demonstrates that there is a gap between the solution directly given by a relaxed ℓ_1 loss and the exact solution required.

We also evaluate the performance on real 3D point cloud road data as in [51]. The task is to determine the 3D points that lie on the road plane (inliers) and those off the plane, *i.e.*, fitting a plane with data dimension $D = 3$. The inlier-outlier threshold is empirically set to *max(0.01, 0.2*noise level)*. Following [51], we use the area under the ROC curve (AUC) to evaluate the performances of each method. We use the data released by the authors of [51], and the average performances of each method are shown in the right plot of Fig. 2. A similar conclusion can be drawn to that of the synthetic data case, the proposed EES method achieves the best result. We also visualize an example of the road plane fitting result in Fig. 3.

Fig. 3: Frame 153 of dataset KITTI-CITY-5: an illustrative example of 3D point cloud road plane fitting, raw image, projection of annotated 3D point cloud onto the image, and detected inliers/outliers using a ground-truth threshold on the distance to the hyperplane for each method. Blue indicates classified inlier points and red indicates the opposite. The noise level is 0.12, the outlier ration is 0.67. The AUC value for EES, EAS and DPCP-PSGM are 0.971, 0.932 and 0.757.

It is also worth discussing the performance gain obtained by EAS compared to DPCP-PSGM. DPCP-PSGM is designed for linear fitting. For linear fitting, EAS adopts the same solver (*i.e.*, projected sub-gradient descent) as in DPCP-PSGM for optimizing the ℓ_1 loss. The difference lies in that our EAS involves an additional post-processing stage as given in Section 3.3. This step can clearly improve the estimation accuracy in low-noise scenarios as shown in the experimental results of Fig. 2.

4.2 Investigation of Different Formulations Based on ℓ_1 Loss

In the study of ℓ_1-based geometric model fitting in Sect. 3.2, several different formulations are introduced to give the relaxed (inexact) solution. This includes four different embeddings, *i.e.* $\mathbf{a}_i$ (3), $\mathbf{b}_i$ (6), $\mathbf{c}_i$ (30) and $\mathbf{d}_i$ (25), and two solvers, *i.e.* hyperplane fitting with Alg. 2 and subspace recovery with Alg. 3. We first study the problem of detecting homography-related correspondences, and leave the discussion for fundamental matrix estimation in the next section.

To detect the homography-related correspondences, possible formulations include $\mathbf{a}_i$ with Alg. 3 to recover a 6-dimensional subspace, and $\mathbf{b}_i$ with Alg. 2 to recover a hyperplane. By also considering the affine approximation for homography, we can leverage $\mathbf{c}_i$ with Alg. 2 to recover

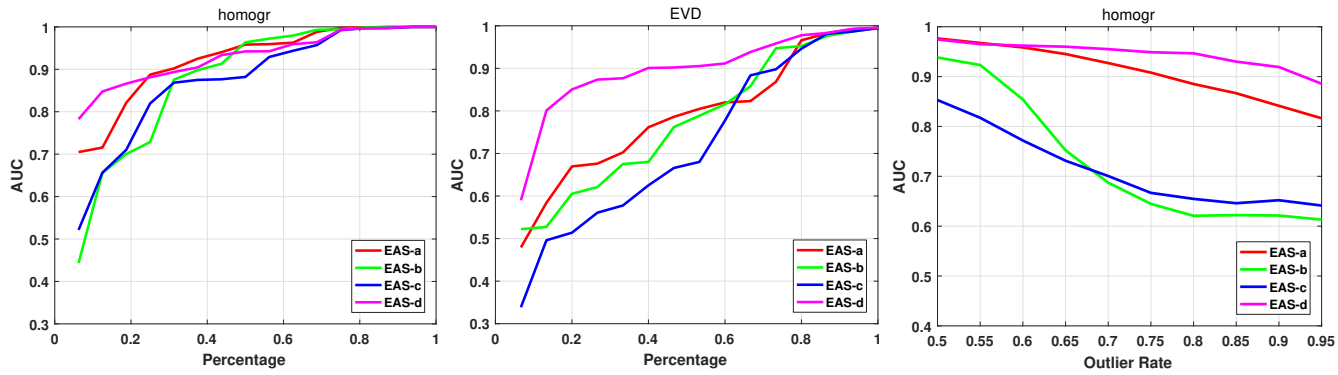


Fig. 4: Evaluation of different formulations for the detection of homography-related correspondences. From left to right, the cumulative distribution of AUC of different formulations on the `homogr` and the `EVD` dataset, and the robustness test against outlier rate with `homogr`. For the cumulative distribution, the better the method performs, the closer its curve is to the top.

a hyperplane, and $\mathbf{d}_i$ with Alg. 3 to recover a 3-dimensional subspace. Next, we provide a thorough experimental study to conclude on the best practice of these variants, as our recommendation.

Note that the detected homography-related correspondences are determined by thresholding the distance of each correspondence to the recovered subspace. With different thresholds, the precision and recall of the detected correspondences *w.r.t.* the ground-truth inliers will vary accordingly. In this sense, the effectiveness of each formulation for an image pair can be evaluated by using the Receiver Operating Characteristic (ROC) curve and Area Under Curve (AUC) statistic.

To conduct the experiments, we adopt the `homogr` and `EVD` datasets that are widely used in the literature for homography estimation [28]. The `homogr` dataset comprises 16 image pairs of relatively short baselines, we use the SIFT [53] algorithm with ratio test at 0.8 to establish the tentative correspondences. Note that for the *CapitalRegion* pair, we fine-tune the threshold to avoid excessive outliers, and for the *LePoint1*, *LePoint2* and *LePoint3* pairs, we use the provided correspondences in the dataset since SIFT fails on these cases. We run RANSAC exhaustively to recover the ground-truth homography transformation, and determine the inliers as the correspondences whose re-projection error are below 3 pixels. The `EVD` dataset is more challenging and comprises of 15 image pairs undergoing extremely view changes. We use the provided tentative correspondence set due to the failure of the SIFT algorithm. The ground-truth homography transformation is provided in this dataset, and we determine the inliers in a similar way as for the `homogr` dataset. For a thorough investigation, we create two testing environments with different settings to evaluate the four formulations. Firstly, the tentative correspondences established by existing algorithms are used for `homogr` and `EVD`, which characterize the general distribution of data. In addition, we also extract the inliers and add a number of outliers by randomly matching two points in the two images to produce correspondence data of a certain outlier rate for the `homogr` dataset. We control the outlier rate in the range of 50% to 95% for the robustness test. For each scene with a given outlier rate, we create 20 instances for the stability of statistics.

We use notations EAS-a, EAS-b, EAS-c and EAS-d to represent the four formulations based on $\mathbf{a}_i$ (3), $\mathbf{b}_i$ (6), $\mathbf{c}_i$ (30) and $\mathbf{d}_i$ (25) for detecting homography-related correspondences, respectively. The results on `homogr` and `EVD` are presented in Fig. 4. We can see that for detecting homography-related correspondences, the formulation with a smaller relative dimension, *i.e.* EAS-a and EAS-d, generally have much better performances. Also, the affine approximation seems to work particularly well and outperforms the counterpart to directly detecting homography-related correspondences. Another observation is that EAS-a and EAS-d are very robust to outlier rate, even in the presence of 95% outliers. This verifies the theoretical results that the ℓ_1-based formulation can tolerate $O(m^2)$ outliers, where m denotes the inlier number. Some representative scenes are presented in Fig. 5, where we show the inliers determined by the threshold of 0.15 for each formulation. We can see that EAS-d produces the correspondences with the best quality. To conclude, EAS-d is the best choice to detect homogarphy-related correspondences, despite its approximation nature.

4.3 Qualitative Comparison of Exact Solution and Relaxed Solution

As we have discussed, in the task of geometric model fitting in the presence of outliers, an exact solution can be achieved by using a truncated loss, and relaxed solution can be achieved by using ℓ_1 loss. However, it is now unclear how the two approaches perform in practice. We next provide an experimental comparison. For the representative image pairs, we adopt four publicly available datasets, *i.e.* `homogr` and `EVD` for homography estimation, `kusvod2` and `AdelaideRMF` for fundamental matrix estimation [28]. The datasets `homogr` and `EVD` have been introduced and the settings here are identical. The dataset `kusvod2` consists of 16 image pairs of both weak and strong perspectives, and we use the provided tentative correspondences in the dataset. The `AdelaideRMF` dataset includes a set of image pairs of campus buildings equipped with manually labelled keypoint correspondences, and we use a 19-pair subset with static scenes. The image pairs are generally of weak perspective since the camera is distant to the building. All the four datasets are provided with a number of annotated ground-truth correspondences, and we use them for evaluation.

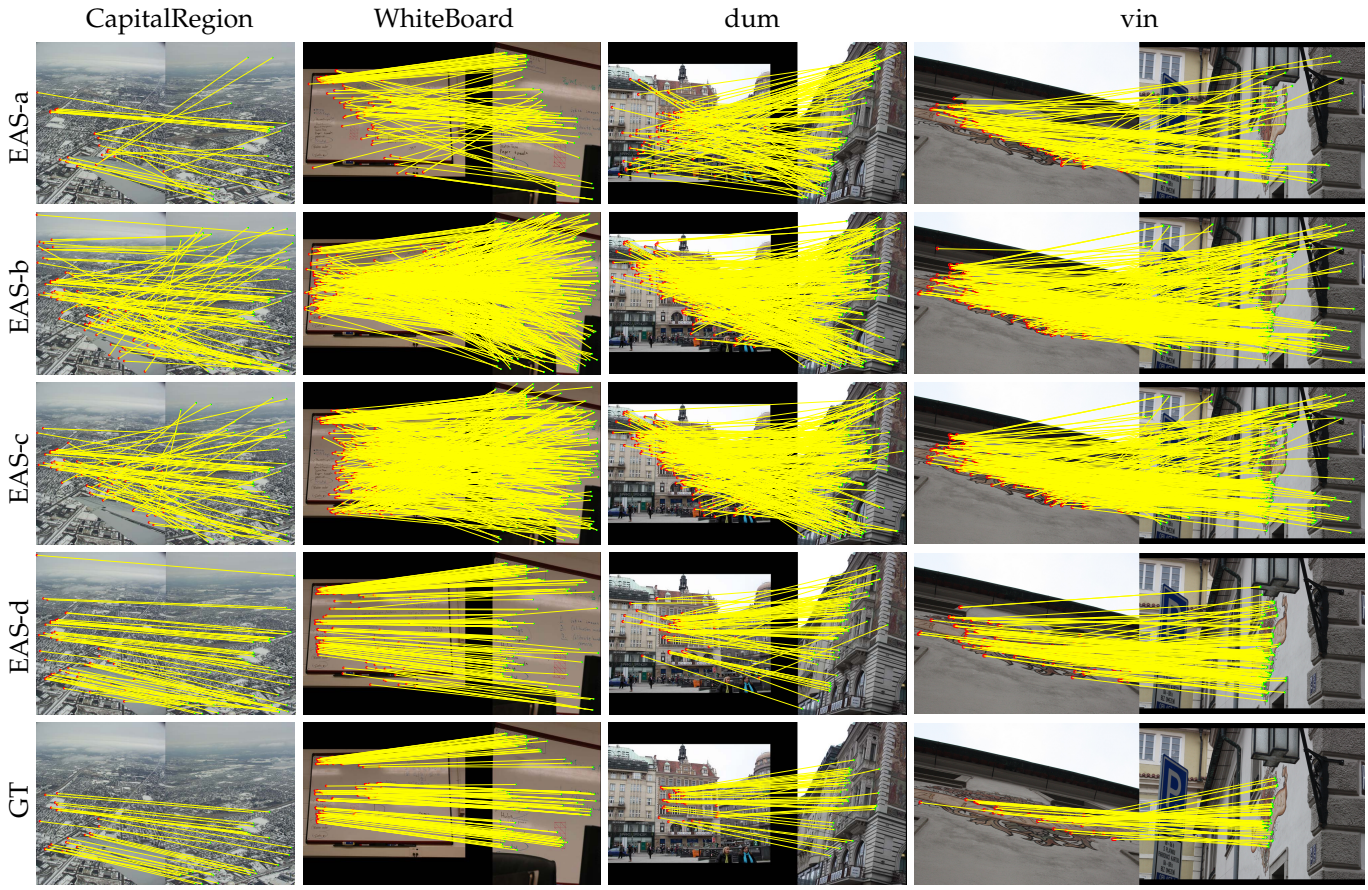


Fig. 5: Representative examples for homography-related correspondences detection. From top to bottom, the correspondences detected by EAS-a, EAS-b, EAS-c, EAS-d and the ground-truth inliers. From left to right, *CapitalRegion* and *WhiteBoard* from `homogr`, and *dum* and *vin* from `EVD`.

Specifically, the performance of each algorithm is evaluated by the average geometric error (Sampson distance in our experiments) of the recovered geometric model *w.r.t.* the annotated ground-truth correspondences.

The proposed method for the exact solution is denoted as EES as aforementioned. For comparison, we adopt the state-of-the-art IBCO method [13], which gives an exact solution in the principle of consensus maximization. As a baseline, and to demonstrate the proposed EES, we additionally include a gradient-descent based solver for truncated ℓ_2 loss, *i.e.* updating the indicator variables using Eq. (17) instead of Eq. (18), which is denoted as GD in our experiments. For relaxed solution with ℓ_1 loss, two strategies are used. **1)** The first is to use $\mathbf{a}_i$ (3), which is a straightforward formulation for both homography and fundamental matrix estimation. We denote this strategy as EAS-F. **2)** The second is based on the observations in Sect. 4.2 that homography-related correspondences can be (approximately) detected leveraging EAS-d. For homography estimation, we can simply use EAS-d with a post-processing stage. For fundamental matrix estimation, we iteratively detect two groups of affine-related correspondences to avoid degeneracy with EAS-d. We denote this strategy as EAS-A.

The experimental results of the five methods, *i.e.* EES, IBCO, GD, EAS-F and EAS-A are presented in Fig. 6. For each image pair, we repeatedly run each method 100 times for the reported statistics. We can observe that the performance of the proposed EES is comparable to the state-of-the-art IBCO, while being orders of magnitude faster. Both methods have significantly outperformed the naive GD. However, it can be seen that the relaxed solution EAS-F and EAS-A with a post-processing stage are substantially more robust than the exact solutions EES and IBCO. It indicates that the exact solution suffers from the issue of non-convexity in optimization, especially in the case of fundamental matrix estimation where many degenerated solutions exist. In conclusion, EES is only suitable for time-critical scenarios with "easy" data for fundamental matrix and homography transformation estimation. In a general scenario, EAS is recommended. As can be seen in Fig. 6, EAS-A outperforms EAS-F in homography estimation, and is marginally better than EAS-F for fundamental matrix estimation. Thus, EAS-A is the most robust one among the proposed methods for geometric model fitting. Some representative examples of EAS-A in detecting affine-related correspondences and the post-processing stage for homography and fundamental matrix estimation are presented in Fig. 7 and Fig. 8, respectively. Also, a quantitative comparison including EAS-F and EAS-A will be provided next.

For our EES which involves the deterministic annealing strategy, we also provide an ablation study regarding the hyper-parameters in it, *i.e.* α_0 and γ. We use the combination of all the four datasets, namely `homogr`, `EVD`, `kusvod2` and `AdelaideRMF` for the experiment. We have tested different

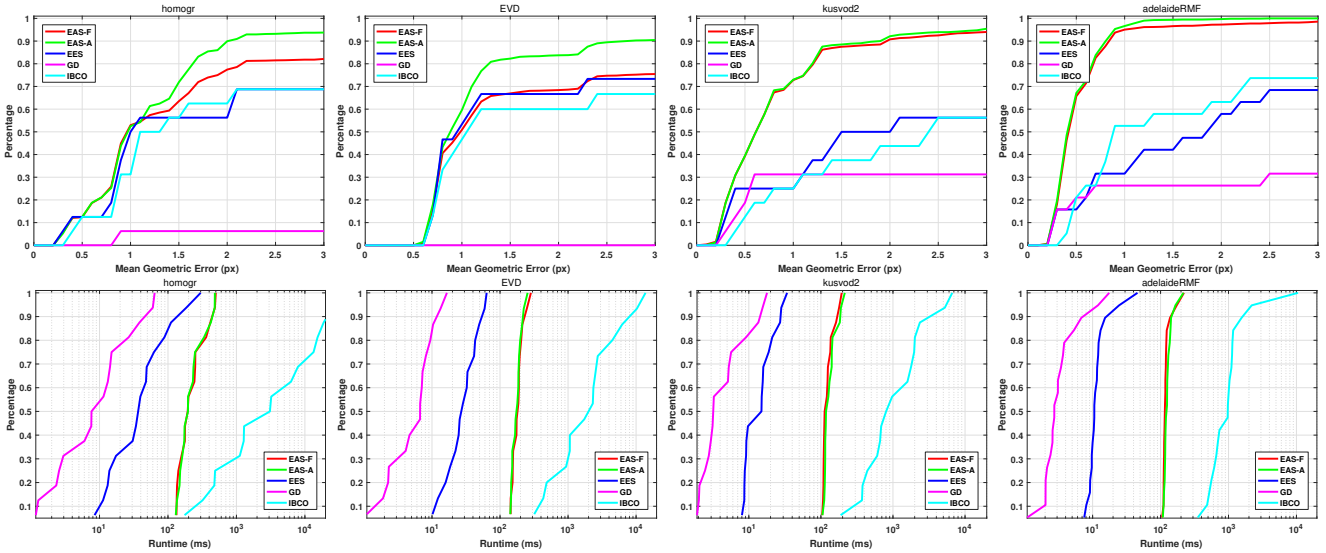


Fig. 6: A qualitative comparison of EAS-F, EAS-A, EES, GD and IBCO on `homogr`, `EVD`, `kusvod2` and `AdelaideRMF`. The first row presents the cumulative distribution of each method *w.r.t.* the mean geometric error, and the second row presents the cumulative distribution of each method *w.r.t.* the runtime. The better the method performs, the closer its curve is to the top.

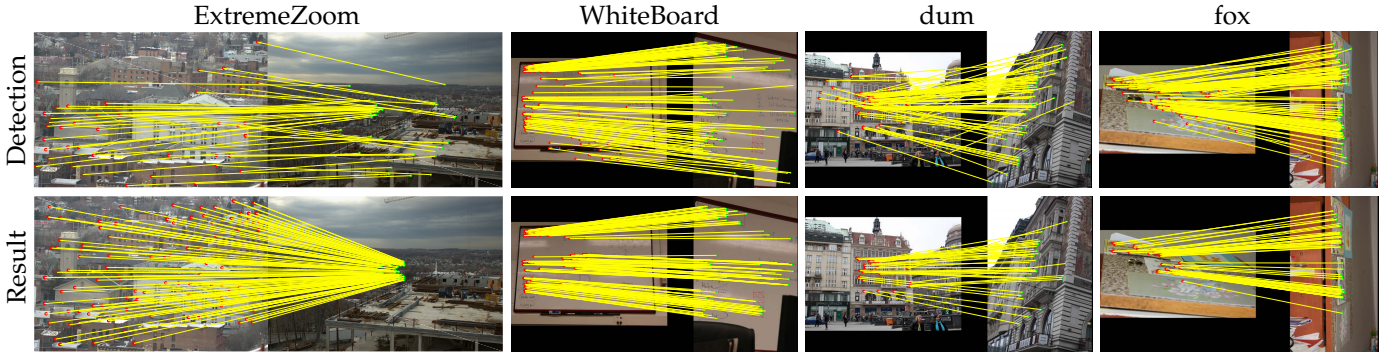


Fig. 7: Some representative examples of EAS-A for homography estimation. The first row presents the correspondences detected by EAS-A, and the second row presents the refined results after the post-processing stage. The adopt image pairs are *ExtremeZoom*, *WhiteBoard* from `homogr`, and *dum* and *fox* from `EVD`.

settings of the hyper-parameters α_0 and γ and the performance of EES, indicated by the cumulative distribution of the geometric error, are presented in Fig. 9. From the results, we can see that the deterministic annealing strategy is quite sensitive to the value of γ, and a large value such as $\gamma = 0.9$ is recommended. This is consistent with the intuition that "slowing down" the annealing process would lead to better resistance against non-convexity. Meanwhile, the strategy is less sensitive to the value of α_0, setting $\alpha_0 = 100$ is recommended and a larger value to slow down the annealing process would not lead to a gain in performance. In conclusion, setting $\alpha_0 = 100$ and $\gamma = 0.9$ is appropriate.

4.4 Quantitative Comparison of EAS and the State-of-The-Art Robust Estimators

To demonstrate the efficacy of our methods, we next provide a quantitative comparison between the proposed EAS-F and EAS-A against the state-of-the-art robust estimators. For homography estimation, we use EAS-A that has been shown to have best performance in previous sections. For fundamental matrix estimation, both EAS-F and EAS-A are included for comparison. As robust estimators, USAC [26], MAGSAC++ [29] (abbreviated as MAG++) and GMS [38] are adopted as the state-of-the-art competitors.

4.4.1 Homography Estimation

The quantitative comparison for homography estimation is conducted on the well-known *HPatches* benchmark [54]. The dataset contains 116 scenes with 696 unique pictures, where the first 57 scenes exhibit illumination changes and the other 59 scenes involve viewpoint changes. Each scene has one reference image, and five target images of varying degrees of illumination or viewpoint changes. For each target image, the ground-truth homography transformation is provided. By matching the reference images with each of the target images, 580 pairs are created for evaluation.

To establish the correspondences, we use SIFT [53] to detect feature points and HardNet [55] to generate descriptors. Then the correspondences are established by nearest-neighbor matching. As for competitors, for USAC, the confidence value is set to 0.99, the maximum number of iterations is set to $10,000$. Note that the pre-defined inlier-outlier

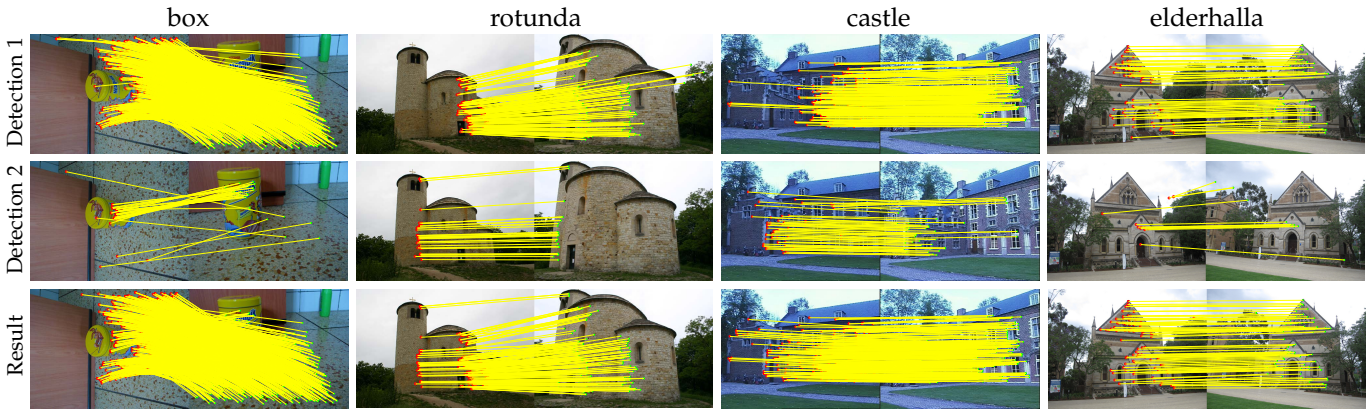


Fig. 8: Some representative examples of EAS-A for fundamental matrix estimation. The first and second row present the correspondences detected by EAS-A in two iterations, and the third row presents the refined results after the post-processing stage. The adopted image pairs are *box*, *rotunda* and *castle* from `kusvod2`, and *elderhalla* from `AdelaideRMF`.

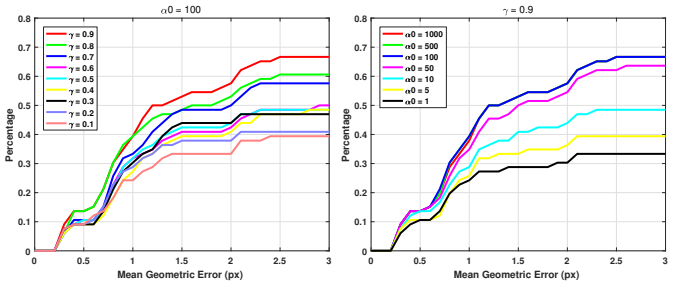


Fig. 9: The ablation experiment for the hyper-parameters introduced in the deterministic annealing strategy, *i.e.* α_0 and γ. The performance is indicated by the cumulative distribution of the geometric error in the combination of the four datasets `homogr`, `EVD`, `kusvod2` and `AdelaideRMF`. The better the method performs, the closer its curve is to the top.

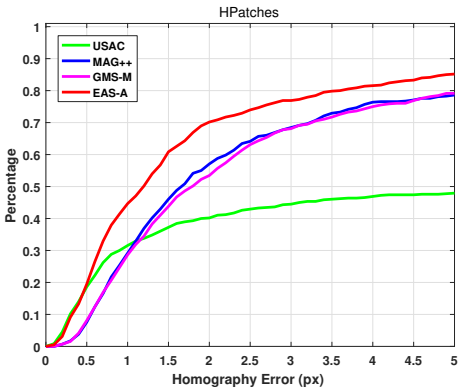


Fig. 10: The cumulative distribution functions of *Homography error* (horizontal axis) of the estimated homography transformation on the *HPatches* dataset. The more accurate the method is, the closer its curve is to the top.

threshold is critical for the performance of USAC, we fine-tune the parameter and set it to 2 pixels. The same inlier-outlier threshold setting also applies to our post-processing algorithm. For MAG++, the inlier-outlier threshold is not required. We follow the default settings in the original paper and set a 1-second time budget to avoid excessive time cost. For GMS, since it cannot directly give the estimates of geometric models, we use it in conjunction with MAG++ (with a 1-second time budget), abbreviated as GMS-M.

We evaluate the performance of each method by comparing the estimated homography with the ground-truth homography. It is not straightforward to directly compare the 3×3 matrices, since different entries in the matrix have different scales. To define a geometrically meaningful metric, we follow SuperPoint [56] to define the *homography error*. In particular, we define the four corners of the reference image as $\mathbf{cn}_1, \mathbf{cn}_2, \mathbf{cn}_3, \mathbf{cn}_4$, and transform them using estimated and ground-truth homography, resulting in $\mathbf{cn}'_1, \mathbf{cn}'_2, \mathbf{cn}'_3, \mathbf{cn}'_4$ and $\hat{\mathbf{cn}}'_1, \hat{\mathbf{cn}}'_2, \hat{\mathbf{cn}}'_3, \hat{\mathbf{cn}}'_4$, respectively. The *homography error* is then defined as the mean value of the distances between $\mathbf{cn}'_i$ and $\hat{\mathbf{cn}}'_i$.

The quantitative comparison results are presented in Fig. 10, and some qualitative examples from the *HPatches* dataset are presented in Fig. 11. From Fig. 10, we can observe that our EAS-A achieves the best accuracy in homography estimation, outperforming USAC, MAGSAC++ and GMS-MAGSAC++ by a large margin. This is because a large portion of outliers exist in the matching process due to large illumination changes or viewpoint changes. Owing to the same reason, GMS fails to boost the performance of MAGSAC++. The superiority of our method can be seen in the visual results in Fig. 11, where the correspondences that are consistent with the estimated homography transformation are shown.

4.4.2 *Fundamental Matrix Estimation*

The quantitative evaluation of fundamental matrix estimation is conducted on the recently introduced benchmark of [57]. Four datasets are introduced in the benchmark, including (i) The `TUM` SLAM dataset [58]: It is of indoor scenes and contains short-baseline image pairs in the resolution of 480×640. (ii) The `KITTI` odometry dataset [59]: It is in a driving scenario, where the geometry between images is dominated by the forward motion. It contains short-baseline image pairs in the resolution of $370 \times 1,226$. (iii) The `Tanks and Temples` (`T&T`) dataset [60]: It provides many scans of scenes or objects for image-based reconstruction, and offers wide-baseline pairs for evaluation. The resolution is $1,080 \times 2,048$ or $1,080 \times 1,920$. (iv) The `Community Photo`

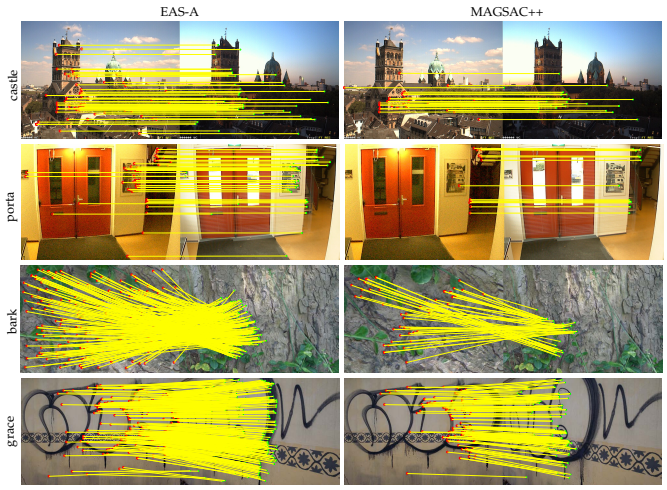


Fig. 11: Some qualitative examples comparing the results of EAS-A (first column) and MAGSAC++ (second column) on *HPatches* dataset. For comparison, we only visualize the correspondences with transfer error below the threshold of 2 pixels.

`Collection` (`CPC`) dataset [61]: It provides unstructured images of well-known landmarks across the world collected from Flickr, taken from arbitrary cameras at a different time. Thus the image pairs are of wide-baseline and the resolution varies.

In our experiment, the SIFT algorithm is used to establish tentative correspondences. In addition, as suggested in [57], the ratio test strategy with threshold as 0.8 is used to pre-prune the potential outliers. For competing algorithms, the settings are retained from the homography estimation experiment, except for USAC, for which we fine-tune the parameter and set it to 1 pixel for the short-baseline dataset `TUM` and `KITTI`, and 2 pixels for the wide-baseline dataset `T&T` and `CPC`.

For all the matchable image pairs, $1,000$ pairs are randomly chosen for each dataset for evaluation. The ground truth is the fundamental matrix for each image pair in the benchmark. In particular, it can be computed from the provided camera intrinsics and extrinsics for `TUM` and `KITTI`. For `T&T` and `CPC`, the ground-truth fundamental matrix is obtained by reconstructing the image sequences using COLMAP [1], which provides accurate estimates of the camera parameters. The accuracy of the estimated fundamental matrix is evaluated by computing the error of [62] referred as *symmetric geometry distance* (SGD) in [57]. Essentially, it is computed by generating virtual correspondences using the ground-truth fundamental matrix and computing the epipolar distance to the estimated one, and then reverting their roles to compute the distance again to ensure symmetry. The averaged distance is then used as SGD. Taking consideration of the different image resolutions, Normalized SGD [57] is more favored in our paper, which is computed as the SGD (in pixels) divided by the length of image diagonals.

The quantitative results are presented in Table 1. We can observe that compared with the results in [57] using plain RANSAC, the state-of-the-art robust estimators USAC and MAG++ can achieve much better results. More interestingly, the performance can be further improved by incorporating a mismatch removal method, such as GMS. It can be seen

TABLE 1: The recall of each method for fundamental matrix estimation on the four datasets. The threshold to determine whether the estimate is accurate or not is 0.05 in terms of Normalized SGD error. Bold indicates the best results.

Dataset	USAC	MAG++	GMS-M	EAS-F	EAS-A
TUM	62.3	70.7	71.3	71.5	**72.8**
KITTI	88.6	88.2	86.9	90.2	**90.4**
T&T	89.1	84.6	92.7	91.3	**92.8**
CPC	60.0	60.1	66.9	67.3	**72.2**

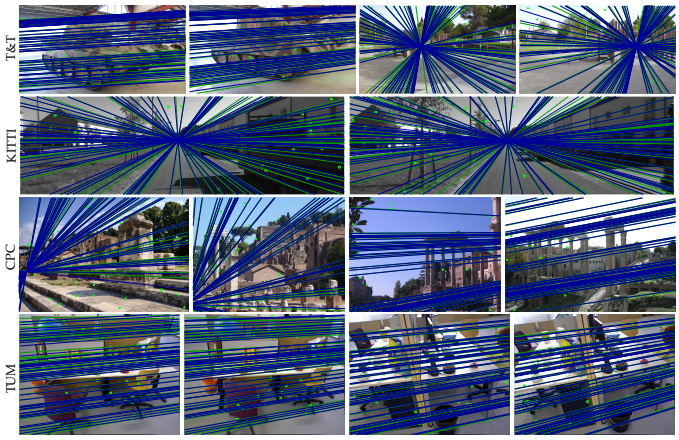


Fig. 12: Qualitative results of EAS-A for fundamental matrix estimation. The blue epipolar lines are produced by the ground-truth fundamental matrix and the green ones are by the estimates. The 1st row represents samples from the dataset of `Tanks and Temples`, the 2nd from `KITTI`, the 3th from `CPC` and the 4th from `TUM`.

that the improvement is quite significant, despite the slight degradation in `KITTI`. As to our methods, it can be seen that EAS-A consistently outperforms the other algorithms, especially on `CPC`, the most challenging wide-baseline dataset. Also, EAS-F is not as robust as EAS-A, which can be seen in the performance comparison on `CPC`. This is because EAS-A induces much simpler optimization problems with smaller relative dimension. Overall, the proposed EAS-A is the most robust and accurate method for geometric estimation, some qualitative results can be seen in Fig. 12.

4.4.3 Additional Results

We also include the comparison results of each method on `homogr`, `EVD`, `kusvod2` and `AdelaideRMF` in Fig. 13. The USAC algorithm is significantly less robust than MAGSAC++, which implies the restriction of requiring an inlier-outlier threshold. The GMS algorithm does not always enhance the estimation accuracy since inliers may also be filtered. It can be seen that our EAS-F and EAS-A achieve the best performances, while EAS-A is the best performer. Note that these datasets are all of wide-baselines, which is consistent to the evaluation result on `CPC`, and clearly demonstrates the advantage of our method in this scenario.

Efficiency of EAS: Due to implementation issues, it is generally hard to directly compare the efficiency of robust estimators with runtime. For example, C++ implementation can be significantly faster than the MATLAB version, and

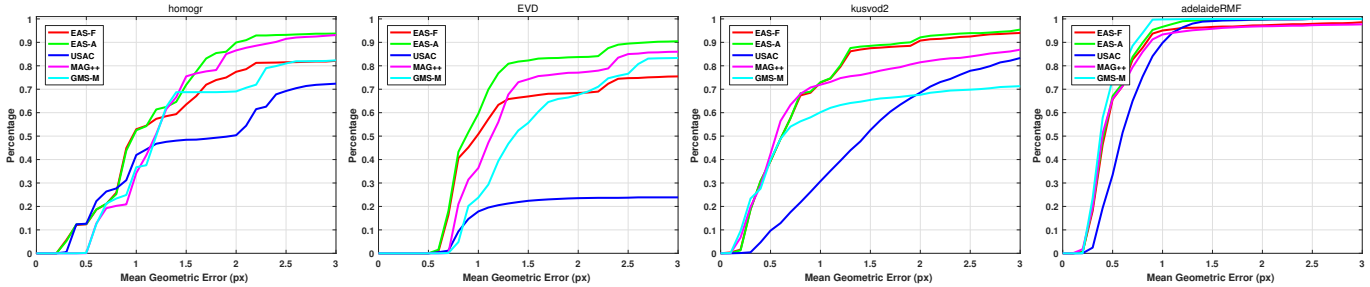


Fig. 13: A comparison of EAS-F, EAS-A, USAC, MAGSAC++ and GMS-MAGSAC++ on `homogr`, `EVD`, `kusvod2` and `AdelaideRMF`. The cumulative distribution of each method *w.r.t.* the mean geometric error are reported. The better the method performs, the closer its curve is to the top.

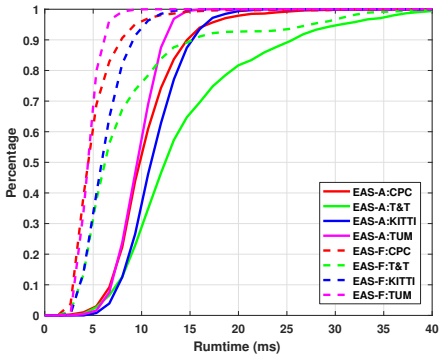


Fig. 14: The cumulative distribution *w.r.t.* runtime of our EAS. Curves of different colors represent different datasets, *i.e.* `TUM`, `KITTI`, `CPC` and `T&T`.

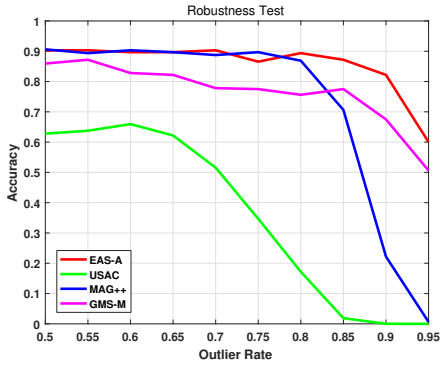


Fig. 15: The robustness test of EAS-A, USAC, MAGSAC++ and GMS-MAGSAC++ on `homogr` dataset.

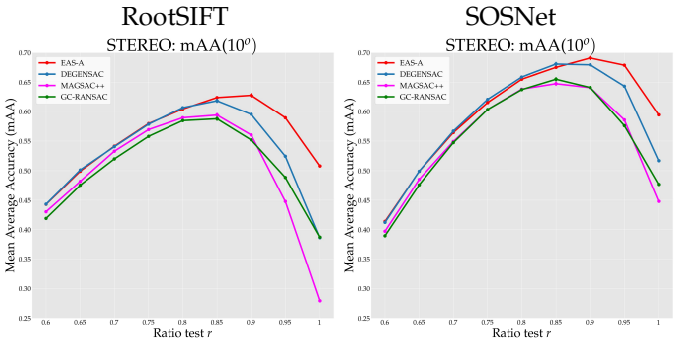


Fig. 16: The comparison of three state-of-the-art robust estimators and EAS-A on the image matching benchmark. The left figure represents the result using RootSIFT, and the right figure represents the result using SOSNet. We report the statistics *w.r.t.* different initial matches determined by the ratio test parameter r. The performance is evaluated in mean Average Accuracy for the estimation of camera pose.

some standard techniques in random sampling scheme can also greatly accelerate the algorithm (such as SPRT [20]). Thus we demonstrate the efficiency of our EAS by decomposing it into two separate parts. Our EAS involves two stages for geometric model fitting, *i.e.* efficient deterministic search and post-processing. Generally the efficient deterministic search stage is very fast using the *projected subgradient-descent* solver. This can be seen in Fig. 14 that our EAS generally require only 20 milliseconds with MATLAB code for most cases to finish this stage. The post-processing stage runs random sampling with 500 iterations, which also admits very efficient implementation within tens of milliseconds. In our implementation, the post-processing stage takes 22.4 ms in average and the mean overall time cost of EAS-A is 42.6 ms, tested on `TUM`, `KITTI`, `CPC` and `T&T`. Thus, in summary, the efficiency of our EAS is quite advantaged, running in tens of milliseconds.

Robustness to Outliers: A nice property of performing deterministic search for geometric model fitting is that it is very insensitive to outliers, in sharp contrast to the traditional random sampling techniques. To verify this point, we conduct a robustness test with the `homogr` dataset, the principle of synthesizing the data with certain outlier rates is previously explained in Sect. 4.2. We use the accuracy, *i.e.* the proportion of instances with mean geometric error below a threshold (*e.g.* 2 pixels in our experiment) to all instances, to evaluate the performance. The results are presented in Fig. 15. It can be seen that our EAS-A and GMS-MAGSAC++ significantly outperform the state-of-the-art random sampling techniques, *i.e.* USAC and MAGSAC++, while our EAS-A achieves the best performance in terms of robustness.

4.5 EAS for Wide-Baseline Stereo

Geometric model fitting is an essential part of modern applications of image matching in computer vision, especially in wide-baseline and multi-view stereo. The recently proposed benchmark [63] has provided a standard pipeline to evaluate the numerous image matching-related techniques that have been proposed, in an extensive and objective way, and focusing on the downstream task of improving the accuracy

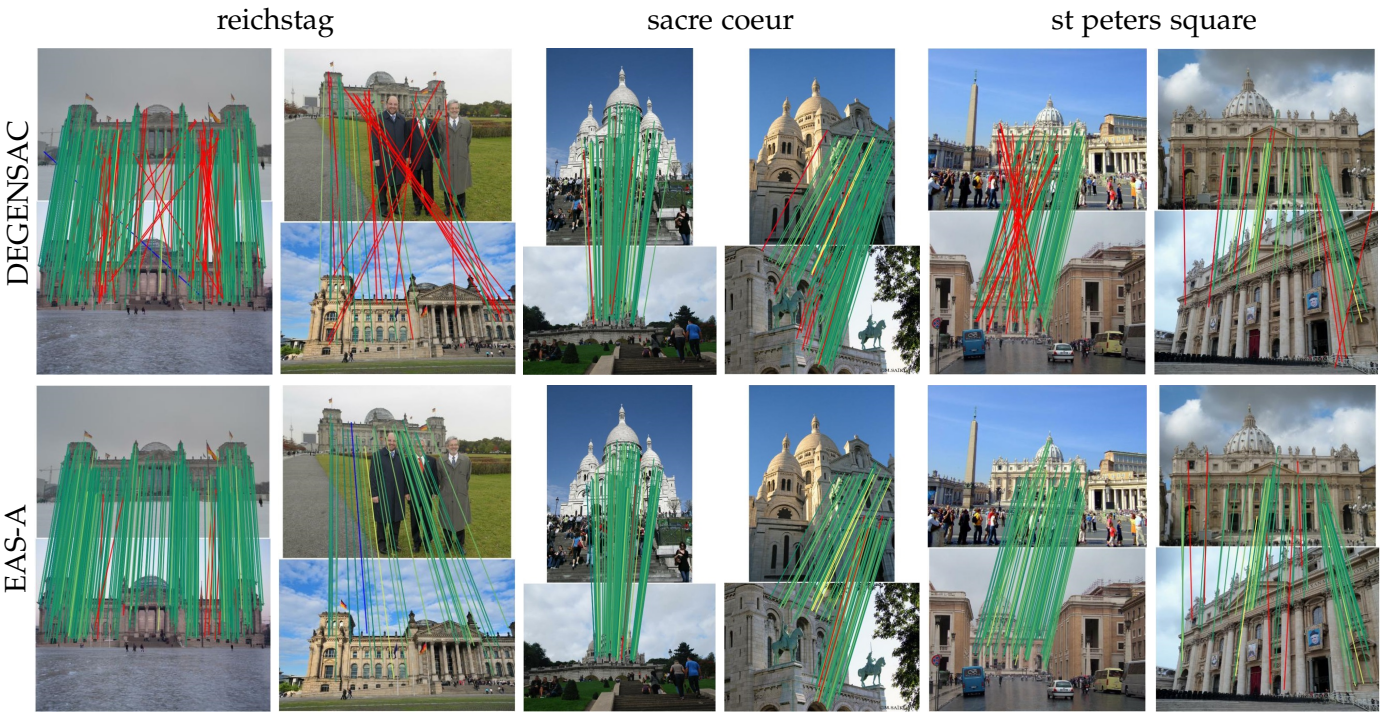


Fig. 17: Some qualitative examples comparing the results of DEGENSAC (first row) and EAS-A (second row). From left to right, the 1st and 2nd columns are from `reichstag`, the 3rd and 4th are from `sacre coeur`, and the 5th and 6th are from `st peters square`. The matches are color-coded by reprojection error computed using the ground-truth depth maps, in green to yellow if they are correct (with green encoding 0 reprojection error and yellow a maximum reprojection error of 5 pixels), in red if they exceed the reprojection error threshold, and in blue if depth estimates are missing.

of the reconstructed camera pose. For the wide-baseline stereo task, the robust estimators, in conjunction with other necessary components, are used to recover the fundamental matrix $\mathbf{F}$ between two images. The known intrinsics $\mathbf{K}$ of the cameras are then used to compute the essential matrix $\mathbf{E} = \mathbf{K}^T\mathbf{F}\mathbf{K}$, which gives the relative rotation and translation vectors with a cheirality check with OpenCV's `recoverPose` function. To evaluate the performance, the main error metric is based on *angular errors* since the stereo problem is defined up to a scale factor [5]. The difference, in degrees, between the estimated and ground-truth translation and rotation vectors between two cameras is computed. Then it is thresholded over a given value for all possible pairs of images. Doing so over different angular thresholds renders a curve, and the mean Average Accuracy (mAA) is computed by integrating this curve up to a maximum threshold, *i.e.* 10° as suggested in [63]. The mAA value is the final error metric. Note that we only evaluate the robust estimators in the wide-baseline stereo track, since in the multi-view stereo track of the benchmark the highly integrated Structure-from-Motion kit COLMAP [1] is used and the robust estimator is not directly replaceable for now in the benchmark.

We use a classical handcrafted method RootSIFT [64] and a state-of-the-art deep-learning based method SOSNet [65] with $8k$ keypoints for constructing the image correspondences. Three robust estimators supported in the benchmark, *i.e.* DEGENSAC [25], GC-RANSAC [27] and MAGSAC++ [29] which are established as the state-of-the-art robust estimators in this task [63], are adopted for comparison. The optimal value 0.5 (pixel) for the inlier-outlier threshold is used for DEGENSAC and GC-RANSAC, and for MAGSAC++, we set the parameter σ_{max} as 5.0 pixels as suggested in the original paper [29]. All compared methods are used with 0.99 confidence and maximum iterations of $10,000$ to balance efficiency and accuracy. In our post-processing stage, the inlier-outlier threshold is also set to 0.5 pixel. The methods are tested on the validation sequences, *i.e.* `reichstag`, `sacre coeur` and `st peters square`. The comparison result of the state-of-the-art robust estimators and our EAS-A is presented in Fig. 16 in terms of mAA. We can observe that generate matching with deep learning based method SOSNet [65] achieves a much better performance than handcrafted RootSIFT [64]. Since large ratio test parameter r generally indicates large outlier rate, it can be seen that owing to the different optimization-based view, our EAS-A is clearly much more robust to outliers than all the other methods, resulting in better mAA both in case of high outlier rates and in terms of a maximum value of all ratio test settings. Some qualitative examples using SOSNet as descriptor can be seen in Fig. 17 with comparison to DEGENSAC, the second-best performer.

5 Conclusion

In this paper, we investigate a long-standing and fundamentally important problem in computer vision, *i.e.* geometric model fitting. The crucial concept in this field, *i.e.* robust loss functions, is re-considered to guide method design, by leveraging the function properties such as exactness and convexity. Based on the inspiration, a group of algorithms are proposed to perform geometric model fitting accurately

and efficiently. Experimental results on publicly available datasets and benchmarks have extensively demonstrated the superiority of our methods, which prevail over the state-of-the-art robust estimators with better robustness and accuracy.

REFERENCES

[1] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4104–4113.

[2] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: a versatile and accurate monocular slam system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[3] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 59–73, 2007.

[4] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *International Journal of Computer Vision*, pp. 1–57, 2020.

[5] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

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

[7] O. Chum, J. Matas, and J. Kittler, "Locally optimized ransac," in *Proceedings of the Joint Pattern Recognition Symposium*, 2003, pp. 236–243.

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

[9] J.-C. Bazin, H. Li, I. S. Kweon, C. Demonceaux, P. Vasseur, and K. Ikeuchi, "A branch-and-bound approach to correspondence and grouping problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1565–1576, 2012.

[10] H. Li, "Consensus set maximization with guaranteed global optimality for robust geometry estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2009, pp. 1074–1080.

[11] T.-J. Chin, P. Purkait, A. Eriksson, and D. Suter, "Efficient globally optimal consensus maximisation with tree search," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2413–2421.

[12] D. Campbell, L. Petersson, L. Kneip, and H. Li, "Globally-optimal inlier set maximisation for simultaneous camera pose and feature correspondence," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1–10.

[13] Z. Cai, T.-J. Chin, H. Le, and D. Suter, "Deterministic consensus maximization with biconvex programming," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 685–700.

[14] H. M. Le, T.-J. Chin, A. Eriksson, T.-T. Do, and D. Suter, "Deterministic approximate methods for maximum consensus robust fitting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[15] K. Aftab and R. Hartley, "Convergence of iteratively re-weighted least squares to robust m-estimators," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 480–487.

[16] A. Fan, X. Jiang, Y. Wang, J. Jiang, and J. Ma, "Geometric estimation via robust subspace recovery," in *Proceedings of the European Conference on Computer Vision*, 2020, pp. 462–478.

[17] P. H. Torr, S. J. Nasuto, and J. M. Bishop, "Napsac: High noise, high dimensional robust estimation-it's in the bag," in *Proceedings of the British Machine Vision Conference*, 2002.

[18] K. Ni, H. Jin, and F. Dellaert, "Groupsac: Efficient consensus in the presence of groupings," in *Proceedings of the IEEE International Conference on Computer Vision*, 2009, pp. 2193–2200.

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

[20] O. Chum and J. Matas, "Optimal randomized ransac," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1472–1482, 2008.

[21] J. Matas and O. Chum, "Randomized ransac with td, d test," *Image and Vision Computing*, vol. 22, no. 10, pp. 837–842, 2004.

[22] P. H. Torr and A. Zisserman, "Mlesac: A new robust estimator with application to estimating image geometry," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.

[23] P. H. S. Torr, "Bayesian model estimation and selection for epipolar geometry and generic manifold fitting," *International Journal of Computer Vision*, vol. 50, no. 1, pp. 35–61, 2002.

[24] K. Lebeda, J. Matas, and O. Chum, "Fixing the locally optimized ransac–full experimental evaluation," in *Proceedings of the British Machine Vision Conference*, 2012, pp. 1–11.

[25] O. Chum, T. Werner, and J. Matas, "Two-view geometry estimation unaffected by a dominant plane," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 772–779.

[26] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J.-M. Frahm, "Usac: a universal framework for random sample consensus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 2022–2038, 2012.

[27] D. Barath and J. Matas, "Graph-cut ransac," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6733–6741.

[28] D. Barath, J. Matas, and J. Noskova, "Magsac: marginalizing sample consensus," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 197–10 205.

[29] D. Barath, J. Noskova, M. Ivashechkin, and J. Matas, "Magsac++, a fast, reliable and accurate robust estimator," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1304–1312.

[30] Y. Zheng, S. Sugimoto, and M. Okutomi, "Deterministically maximizing feasible subsystem for robust model fitting with unit norm constraint," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1825–1832.

[31] Z. Cai, T.-J. Chin, and V. Koltun, "Consensus maximization tree search revisited," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 1637–1645.

[32] C. Olsson, O. Enqvist, and F. Kahl, "A polynomial-time bound for matching and registration with outliers," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[33] O. Enqvist, E. Ask, F. Kahl, and K. Åström, "Robust fitting for multiple view geometry," in *Proceedings of the European Conference on Computer Vision*, 2012, pp. 738–751.

[34] P. Purkait, C. Zach, and A. Eriksson, "Maximum consensus parameter estimation by reweighted ℓ_1 methods," in *Proceedings of the International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2017, pp. 312–327.

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

[36] 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 34–47, 2017.

[37] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *International Journal of Computer Vision*, vol. 127, no. 5, pp. 512–531, 2019.

[38] J.-W. Bian, W.-Y. Lin, Y. Liu, L. Zhang, S.-K. Yeung, M.-M. Cheng, and I. Reid, "GMS: Grid-based motion statistics for fast, ultra-robust feature correspondence," *International Journal of Computer Vision*, vol. 128, no. 6, pp. 1580–1593, 2020.

[39] K. M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, and P. Fua, "Learning to find good correspondences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2666–2674.

[40] J. Zhang, D. Sun, Z. Luo, A. Yao, L. Zhou, T. Shen, Y. Chen, L. Quan, and H. Liao, "Learning two-view correspondences and geometry using order-aware network," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 5845–5854.

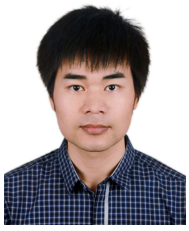
[41] W. Sun, W. Jiang, E. Trulls, A. Tagliasacchi, and K. M. Yi, "Acne: Attentive context normalization for robust permutation-equivariant learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11 286–11 295.

[42] C. Choy, J. Lee, R. Ranftl, J. Park, and V. Koltun, "High-dimensional convolutional networks for geometric pattern recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11 227–11 236.

[43] M. C. Tsakiris and R. Vidal, "Dual principal component pursuit," *Journal of Machine Learning Research*, vol. 19, no. 1, pp. 684–732, 2018.

[44] G. Lerman and T. Maunu, "Fast, robust and non-convex subspace recovery," *Information and Inference: A Journal of the IMA*, vol. 7, no. 2, pp. 277–336, 2018.
[45] T. Maunu, T. Zhang, and G. Lerman, "A well-tempered landscape for non-convex robust subspace recovery." *J. Mach. Learn. Res.*, vol. 20, no. 37, pp. 1–59, 2019.
[46] G. Lerman, M. B. McCoy, J. A. Tropp, and T. Zhang, "Robust computation of linear models by convex relaxation," *Foundations of Computational Mathematics*, vol. 15, no. 2, pp. 363–410, 2015.
[47] H. Xu, C. Caramanis, and S. Sanghavi, "Robust pca via outlier pursuit," in *Advances in neural information processing systems*, 2010, pp. 2496–2504.
[48] G. Lerman and T. Maunu, "An overview of robust subspace recovery," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1380–1410, 2018.
[49] T. Ding, Y. Yang, Z. Zhu, D. P. Robinson, R. Vidal, L. Kneip, and M. C. Tsakiris, "Robust homography estimation via dual principal component pursuit," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6080–6089.
[50] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2210–2239, 1998.
[51] Z. Zhu, Y. Wang, D. Robinson, D. Naiman, R. Vidal, and M. Tsakiris, "Dual principal component pursuit: Improved analysis and efficient algorithms," in *Advances in Neural Information Processing Systems*, 2018, pp. 2171–2181.
[52] T. Ding, Z. Zhu, T. Ding, Y. Yang, D. Robinson, R. Vidal, and M. Tsakiris, "Noisy dual principal component pursuit," in *Proceedings of the International Conference on Machine Learning*, 2019, pp. 1617–1625.
[53] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
[54] V. Balntas, K. Lenc, A. Vedaldi, and K. Mikolajczyk, "Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5173–5182.
[55] A. Mishchuk, D. Mishkin, F. Radenovic, and J. Matas, "Working hard to know your neighbor's margins: Local descriptor learning loss," in *Advances in Neural Information Processing Systems*, 2017, pp. 4829–4840.
[56] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 224–236.
[57] J.-W. Bian, Y.-H. Wu, J. Zhao, Y. Liu, L. Zhang, M.-M. Cheng, and I. Reid, "An evaluation of feature matchers for fundamental matrix estimation," in *Proceedings of the British Machine Vision Conference*, 2019.
[58] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 573–580.
[59] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 3354–3361.
[60] A. Knapitsch, J. Park, Q.-Y. Zhou, and V. Koltun, "Tanks and temples: Benchmarking large-scale scene reconstruction," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1–13, 2017.
[61] K. Wilson and N. Snavely, "Robust global translations with 1dsfm," in *Proceedings of the European Conference on Computer Vision*, 2014, pp. 61–75.
[62] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *Int. J. Comput. Vis.*, vol. 27, no. 2, pp. 161–195, 1998.
[63] Y. Jin, D. Mishkin, A. Mishchuk, J. Matas, P. Fua, K. M. Yi, and E. Trulls, "Image matching across wide baselines: From paper to practice," *International Journal of Computer Vision*, vol. 129, no. 2, pp. 517–547, 2021.
[64] R. Arandjelović and A. Zisserman, "Three things everyone should know to improve object retrieval," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2911–2918.
[65] Y. Tian, X. Yu, B. Fan, F. Wu, H. Heijnen, and V. Balntas, "Sosnet: Second order similarity regularization for local descriptor learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 016–11 025.

Aoxiang Fan received the B.S. degree from the Electronic Information School,Wuhan University, Wuhan, China, in 2018. He is currently pursuing the master's degree with the Multi-Spectral Vision Processing Lab, Wuhan University. His current research interests include computer vision and pattern recognition.



Jiayi Ma received the B.S. degree in information and computing science and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. He is currently a Professor with the Electronic Information School, Wuhan University. He has authored or co-authored more than 200 refereed journal and conference papers, including IEEE TPAMI/TIP, IJCV, CVPR, ICCV, ECCV, *etc.* His research interests include computer vision, machine learning, and pattern recognition. Dr. Ma has been identified in the 2019 and 2020 Highly Cited Researcher lists from the Web of Science Group. He is an Area Editor of *Information Fusion*, an Associate Editor of *Neurocomputing*, and a Guest Editor of *Remote Sensing*.

Xingyu Jiang received the B.E. degree from the Department of Mechanical and Electronic Engineering, Huazhong Agricultural University, Wuhan, China, in 2017, and the M.S. degree from the Electronic Information School, Wuhan University, Wuhan, China, in 2019. He is currently a Ph.D. student with the Electronic Information School, Wuhan University. His research interests include computer vision, machine learning, and pattern recognition.

Haibin Ling received the B.S. and M.S. degrees from Peking University in 1997 and 2000, respectively, and the Ph.D. degree from the University of Maryland, College Park, in 2006. From 2000 to 2001, he was an assistant researcher at Microsoft Research Asia. From 2006 to 2007, he worked as a postdoctoral scientist at the University of California Los Angeles. In 2007, he joined Siemens Corporate Research as a research scientist; then, from 2008 to 2019, he worked as a faculty member of the Department of Computer Sciences at Temple University. In fall 2019, he joined Stony Brook University as a SUNY Empire Innovation Professor in the Department of Computer Science. His research interests include computer vision, augmented reality, medical image analysis, and human computer interaction. He received Best Student Paper Award at ACM UIST (2003), NSF CAREER Award (2014), Yahoo Faculty Research Award (2019), and Amazon AWS Machine Learning Research Award (2019). He serves as Associate Editors for several journals including IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), Pattern Recognition (PR), and Computer Vision and Image Understanding (CVIU), and has served as Area Chairs various times for CVPR and ECCV.