Geometric Estimation via Robust Subspace Recovery

Aoxiang Fan[1], Xingyu Jiang[1], Yang Wang[1], Junjun Jiang[2,3], and Jiayi Ma[1(✉)]

[1] Wuhan University, Wuhan 430072, China
{fanaoxiang,jiangx.y,wangyang}@whu.edu.cn,
jyma2010@gmail.com
[2] Harbin Institute of Technology, Harbin 150001, China
jiangjunjun@hit.edu.cn
[3] Peng Cheng Laboratory, Shenzhen 518000, China

Abstract. Geometric estimation from image point correspondences is the core procedure of many 3D vision problems, which is prevalently accomplished by random sampling techniques. In this paper, we consider the problem from an optimization perspective, to exploit the intrinsic linear structure of point correspondences to assist estimation. We generalize the conventional method to a robust one and extend the previous analysis for linear structure to develop several new algorithms. The proposed solutions essentially address the estimation problem by solving a subspace recovery problem to identify the inliers. Experiments on real-world image datasets for both fundamental matrix and homography estimation demonstrate the superiority of our method over the state-of-the-art in terms of both robustness and accuracy.

Keywords: Geometric estimation · Robust model fitting · 3D vision · Robust subspace recovery

1 Introduction

In 3D vision, a vast majority of applications, such as image stitching [5], structure-from-motion [17] and simultaneous localization and mapping [26], rely on feature point correspondences between images for geometric estimation. However, due to the imperfections of both local key point detection and feature description techniques, the correspondences are invariably contaminated by noises and a number of outliers. The degenerated data pose great challenges for accurate estimation. Consequently, the most widely used estimator in practice has become the well-known random sample consensus (RANSAC) [15], despite its simplicity and time of invention.

In essence, RANSAC proceeds by repeatedly sampling a minimal subset of correspondences to propose hypothesis, *e.g.*, 4 correspondences for homography

J. Ma—This work was supported by the National Natural Science Foundation of China under Grant nos. 61773295 and 61971165.

A. Vedaldi et al. (Eds.): ECCV 2020, LNCS 12367, pp. 462–478, 2020.
https://doi.org/10.1007/978-3-030-58542-6_28

and 7 for fundamental estimation. The process is iterated until a convergence criterion that provides a probability guarantee of hitting an all-inlier subset is met. However, some fundamental shortcomings exist with the randomized hypothesize-and-verify search strategy. One of the main limitations lies in the degraded performance against dominant outliers. The required time to retrieve an all-inlier subset grows exponentially with respect to the outlier ratio, and the estimation accuracy also suffers from high variance.

Targeting on more accurate estimation results and minimum processing time, almost all phases of the random sampling estimator has been investigated, leading to a large group of RANSAC variants. For acceleration, one particularly successful strategy is to incorporate additional prior information, such as matching score [9] and spatial coherence of correspondences [24,27] to increase the probability of hitting an all-inlier subset. The verification stage has also been modified to avoid unnecessary computations [10,25]. However, due to the randomized nature, efficiency and robustness (or accuracy) are often contradictory to each other. From an optimization perspective, there exists a group of methods known as *consensus maximization* in the literature for accurate geometric estimation [3,4,18,22]. In contrast to random sampling techniques, these methods translate the geometric estimation problem into an optimization problem, in which the technical difficulty becomes the highly non-convex objective. Based on different theoretical guarantees, they can be roughly divided into two categories, global methods and local methods. However, despite the nice property of theoretical optimality, the global methods are generally computationally demanding. While the local methods are faster, but still too time-consuming for real-time applications.

In this paper, we also consider the geometric estimation in an optimization point of view. Differently, we aim at exploring the intrinsic linear structure of point correspondences. This linear structure is actually previously addressed in the literature, and can be traced to the classic direct linear transformation (DLT) [17]. As DLT reveals, geometric estimation task can be accomplished by solving a (possibly over-determined) set of linear equations derived from the inliers. Considering the noise, DLT finds the solution minimizing square error. In practice, DLT is always used in conjunction with random sampling techniques since it cannot handle outliers. In our method, we further explore the direction to make full use of the linear structure. We provide a more in-depth analysis of the linear structure, and excavate it with an outlier-robust ℓ_1-based objective that significantly extends DLT. The resulting optimization problems are shown to be special forms of robust subspace recovery [20], which allows recently developed efficient and theoretically well-grounded methods to be applied.

To conclude, our contributions include three aspects. (i) We propose a novel method with an outlier-robust objective to excavate of the intrinsic linear structure of point correspondences for geometric estimation. The method exhibits better efficiency and preferable accuracy in case of dominant outliers compared to the current state-of-the-art methods. (ii) We generalize the linear structure discussed in conventional methods and propose several new algorithms based

on robust subspace recovery to take advantage of it. (iii) We demonstrate our method on real-world images for both fundamental and homography estimation tasks with comparisons to current state-of-the-art methods.

2 Related Work

There is a large volume of methods in the literature proposed to address the geometric estimation 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 robustness for geometric estimation, 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 plain RANSAC, 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 [31] and GroupSAC [27], and matching scores in EVSAC [16] and PROSAC [9]. Moreover, improving the model verification stage has also been shown to be critical for the efficiency concern [10,25]. There are also some efforts that have proven to be able to obtain more accurate estimation results. These methods include MLESAC [32] and MSAC [30], in which the model quality is evaluated with a maximum likelihood process. A more illuminating idea is proposed in locally optimized RANSAC (LO-RANSAC) [11,19], where a local optimization step is introduced to polish the so-far-the-best model. By involving more inliers for estimation, the bias induced of noises is reduced and a more accurate model can be expected. This idea has been recently extended by Graph-cut RANSAC [1]. Notably, by combining the most promising improvements, the USAC [29] is proposed as the state-of-the-art of RANSAC variants.

From a different perspective, the geometric estimation problem can be and has been addressed in an optimization framework, with the concept of consensus maximization. The consensus maximization objective 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. Thus, a variety of methods attempt to develop algorithms to search the solution with global optimality guarantee [3,4,6,8,14,22]. However, the computational complexity of these methods is generally prohibitively high due to the fundamental intractability of the problem. Faster optimization-based methods have recently been developed without global optimality [7,18,28], yet still require excessive time compared to random sampling techniques.

Our method follows a different point of view from both categories of the above efforts. We focus on exploring the intrinsic linear structure that has been utilized for decades in DLT. The DLT algorithm is widely used for geometric estimation with the outlier-free data only contaminated by noise. This is hardly the case in practice and as a result, DLT is always used in conjunction with

random sampling techniques. In addition, as will be demonstrated, the resulting problem for exploring the linear structure is related to robust subspace recovery. Recent advances [21,23,33] in this field have shown great potential in handling more degraded cases, *e.g.*, dominant outliers. We refer the interested readers to the recent survey [20] for a comprehensive understanding.

3 Method

Suppose we are given a set of 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. We aim 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.

3.1 Preliminaries on DLT

We first give a brief review the classic DLT algorithm, which provides an efficient solution for geometric estimation by excavating the intrinsic linear structure in the data. The linear structure of point correspondences indicates that, the geometric model can be estimated by solving a linear system induced by the data. In the following, we will discuss the DLT solution for fundamental matrix $\mathbf{F}$ and homography transformation $\mathbf{H}$, respectively.

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

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

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:

$$\mathbf{x}_i'^T \times \mathbf{H}\mathbf{x}_i = \mathbf{0}. \tag{2}$$

For both the cases of estimating $\mathbf{F}$ and $\mathbf{H}$, it reduces to solving an overdetermined linear system in DLT:

$$\mathbf{M}\mathbf{z} = \mathbf{0}, \tag{3}$$

where $\mathbf{M} \in \mathbb{R}^{K\times D}$ represents the data matrix derived from the correspondences and $\mathbf{z} \in \mathbb{R}^D$ represents the column vector of parameters. For fundamental matrix estimation, we have $\mathbf{M} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]^T$ and the embedding is

$$\mathbf{a}_i = (x'_i x_i, x'_i y_i, x'_i, y'_i x_i, y'_i y_i, y'_i, x_i, y_i, 1)^T. \tag{4}$$

For homography estimation, we have $\mathbf{M} = [\mathbf{b}_1^T, \mathbf{b}_2^T, \ldots, \mathbf{b}_n^T]^T$ and the embedding is

$$\mathbf{b}_i^T = \begin{bmatrix} \mathbf{0}^T & -\mathbf{x}_i^T & y'_i \mathbf{x}_i^T \\ \mathbf{x}_i^T & \mathbf{0}^T & -x'_i \mathbf{x}_i^T \end{bmatrix}. \tag{5}$$

The solution is given as the right singular vector corresponding to the smallest singular value of $\mathbf{M}$.

3.2 Robust Generalization

In an outlier-free scenario, DLT is known to be able to give near-optimal results. *Therefore, it is natural to ask that: can we extend the framework of DLT to cope with outliers?* In this section, we try to give a positive answer.

First, we consider the ideal case for estimation, where there exists no noise for the correspondences. In that case, we can reformulate DLT as follows by taking outliers into consideration:

$$\min_{\mathbf{z}\in\mathbb{R}^D} \quad \|\mathbf{Mz}\|_0, \quad s.t. \quad \mathbf{z} \neq \mathbf{0}. \tag{6}$$

The ℓ_0-based functional $\|\mathbf{Ez}\|_0$ simply computes how many points do not conform to the linear structure. Since no linear structure is expected for outliers, the solution of (6) will in general be the ground truth estimation. However, ℓ_0 minimization is known to be computationally intractable, thus we replace it by the following ℓ_1 minimization problem:

$$\min_{\mathbf{z}\in\mathbb{R}^D} \quad \|\mathbf{Mz}\|_1, \quad s.t. \quad \|\mathbf{z}\|_2 = 1. \tag{7}$$

In this sense, the geometric task becomes numerically solvable, and also, applicable to noise-contaminated data. The relation between (7) and DLT is clear, *i.e.* a DLT solution equals to using an ℓ_2-based objective for (7). This explains the limitation of DLT, since an ℓ_2 objective is known to be sensitive to outliers.

Mathematically, (6) and (7) can be seen as a hyperplane fitting problem. In fact, the exact form of (7) has been recently investigated in the literature of robust subspace recovery [33,35], where hyperplane fitting is a special case when the intrinsic dimension of data $d = D - 1$. The robust property has been theoretically demonstrated, which roughly states that under some assumptions on the distributions of outliers, the estimation task with (7) can even tolerate $O(m^2)$ outliers, where m denotes the inlier number.

Note that (7) 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 (7) have been proposed. In [33], (7) is relaxed to a sequence of *linear programs*, which guarantees finite convergence to the global optima. However, this approach is computationally expensive. Alternatively, [33] 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 [35]. 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. Also, it is important to clarify that due to the relaxation strategy, the (global) solution for (7) may not be ideal under noise. In fact, it has been proven that the global minimizer is perturbed away from the ground truth by an amount proportional to the noise level (while still tolerate $O(m^2)$ outliers) [13]. Thus, the solution is generally coarse and needs to be refined by post-processing, as we will discuss in Sect. 3.4. We outline the proposed geometric estimation method with (7) in Algorithm 1.

Algorithm 1. Geometric Estimation with Hyperplane Fitting

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

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: $\mathbf{g} = \mathbf{M}^T sign(\mathbf{Mz})$.
5: Update the step size μ according to a certain rule.
6: Sub-gradient descent: $\mathbf{z} = \mathbf{z} - \mu\mathbf{g}$.
7: Sphere projection: $\mathbf{z} = \mathbf{z}/\|\mathbf{z}\|_2$.
8: **end while**
9: Compute the residuals of each correspondence with respect to $\mathbf{z}$.
10: Determine the inlier set $\mathcal{I}$ by thresholding the residuals.
11: Post-processing on $\mathcal{I}$ to obtain final estimation result $\mathbf{s}$.
12: **return** $\mathbf{s}$.

3.3 Extended Exploration of Linear Structure

Although the robust formulation in (7) has encouraged several effective algorithms for geometric estimation, there are some critical issues when dealing with real-world corrupted data. Due to the inherent non-convexity of the problem, the locally convergent algorithm can easily be trapped in weak local optima and fail to give meaningful results, especially when the data suffer from strong noise or heavy outliers. These motivate us to reconsider the problem in response to the great challenge posed by practical applications.

Linear Structure with Affine Camera. First, let us start from the simple case of affine camera model and showcase how to exploit its linear structure. If both views are assumed to be taken by an affine camera, the two matched feature points are related by an affine transformation:

$$\mathbf{x}_i' = \mathbf{A}\mathbf{x}_i, \tag{8}$$

where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix} \tag{9}$$

represents the affine matrix.

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:

$$\mathbf{c}_i^T = \begin{bmatrix} \mathbf{x}_i^T & \mathbf{0}^T & -x_i' \\ \mathbf{0}^T & \mathbf{x}_i^T & -y_i' \end{bmatrix}. \tag{10}$$

The problem can be then readily solved using (7) 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.

Note that geometric estimation using (7) actually advocates a subspace recovery problem, with the data corrupted by noise and outliers. In the view of 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 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. To this end, we next present an alternative formulation to exploit the linear structure, with a lower relative dimension to deal with.

Since an affine transformation only involves linear terms with respect to the correspondence data, we consider the following embedding

$$\mathbf{d}_i = [x_i, y_i, x_i', y_i', 1]^T. \tag{11}$$

The structure of this embedding is revealed by the following equation derived from (8):

$$\mathbf{A}'\mathbf{d}_i = 0, \tag{12}$$

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

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

Since $\mathbf{A}'$ is clearly of rank 2, this indicates that 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.

The above observation suggests solving the following 3-dimensional subspace recovery problem:

$$\min_{\mathbf{v} \in \mathbb{R}^{D \times 2}} \quad \textstyle\sum_i \|\mathbf{d}_i^T \mathbf{v}\|_2 = \|\mathbf{M}\mathbf{v}\|_{1,2}, \quad s.t. \;\; \mathbf{v}^T\mathbf{v} = \mathbf{I}, \tag{14}$$

where $\mathbf{M} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n]^T$, $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2]$ represents the matrix of two orthogonal unit vectors, $\mathbf{I}$ represents the identity matrix, and $\|\cdot\|_{1,2}$ represents the sum of the Euclidean norms of the rows of the input matrix. The relative dimension is 3/5 for (14), which is much smaller than 6/7 indicated by the hyperplane fitting case (10). This renders the problem a much easier task for learning.

The rationale behind (14) is to find two 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.* [21], as discussed in the comprehensive survey [20]. In this paper, we adopt a more efficient strategy to iteratively search the two bases. In the first iteration, a hyperplane fitting algorithm is conducted to find the first basis. In the second iteration, the procedure is similar to hyperplane fitting but with an additional projection step to find the second basis. The additional projection step guarantees that the second basis is orthogonal to the first one. Specifically, if we obtain the first basis $\mathbf{v}_1$, the projector of its orthogonal complement should be $\mathbf{I} - \mathbf{v}_1\mathbf{v}_1^T$, then the second basis $\mathbf{v}_2$ should be projected onto it as $\mathbf{v}_2 = (\mathbf{I} - \mathbf{v}_1\mathbf{v}_1^T)\mathbf{v}_2 = \mathbf{v}_2 - \mathbf{v}_1\mathbf{v}_1^T\mathbf{v}_2$. The algorithm to solve (14) is outlined in Algorithm 2.

Algorithm 2. Inlier Detection with Linear Embedding

Input: The correspondence set $\mathcal{S}$
Output: The estimated bases $\mathcal{I}$

Mapping the correspondences into embeddings $\mathbf{d}_i$ to form the data matrix $\mathbf{M}$.
Initialize $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2]$ as the right singular vectors of the two smallest singular values of $\mathbf{M}$.
while not converge **do**
  Compute sub-gradient: $\mathbf{g}_1 = \mathbf{M}^T sign(\mathbf{M}\mathbf{v}_1)$.
  Update the step size μ according to a certain rule.
  Sub-gradient Descent: $\mathbf{v}_1 = \mathbf{v}_1 - \mu_i \mathbf{g}_1$.
  Sphere Projection: $\mathbf{v}_1 = \mathbf{v}_1 / \|\mathbf{v}_1\|_2$.
end while
Orthogonal Projection: $\mathbf{v}_2 = \mathbf{v}_2 - \mathbf{v}_1\mathbf{v}_1^T\mathbf{v}_2$.
Sphere Projection: $\mathbf{v}_2 = \mathbf{v}_2 / \|\mathbf{v}_2\|_2$.
while not converge **do**
  Compute sub-gradient: $\mathbf{g}_2 = \mathbf{M}^T sign(\mathbf{M}\mathbf{v}_2)$.
  Update the step size ν according to a certain rule.
  Sub-gradient Descent: $\mathbf{v}_2 = \mathbf{v}_2 - \nu_j \mathbf{g}_2$.
  Orthogonal Projection: $\mathbf{v}_2 = \mathbf{v}_2 - \mathbf{v}_1\mathbf{v}_1^T\mathbf{v}_2$.
  Sphere Projection: $\mathbf{v}_2 = \mathbf{v}_2 / \|\mathbf{v}_2\|_2$.
end while
Compute the residuals of each correspondence with respect to $\mathbf{v}$.
Determine the inlier set $\mathcal{I}$ by thresholding the residuals.
return $\mathcal{I}$.

Extended Linear Structure. From the discussion above, we can conclude that with an affine camera, the problem admits a much simpler solution exploiting the linear embedding $\mathbf{d}_i$. *Thus, a natural question is, can Algorithm 2 be extended to more general scenarios where the affine camera assumption is not strictly satisfied?* Next, we try to positively answer the question.

To respond to the question, the first step is to answer that, *what can we obtain from Algorithm 2 in general scenarios?* There are different angles to put an analysis to it, as we will explain in detail next.

The first discussion is based on some important conclusions in Sect. 6.3 of [17]. We denote $\mathbf{x}_p$ the measured image feature point from a general finite projective camera, and $\mathbf{x}_a$ the virtual image feature point of the same 3D point, but from the virtual camera at infinity. It can be deduced that $\tilde{\mathbf{x}}_p$ and $\tilde{\mathbf{x}}_a$ are related by the following equation:

$$\tilde{\mathbf{x}}_a - \tilde{\mathbf{x}}_p = \frac{\Delta}{d_0}(\tilde{\mathbf{x}}_p - \tilde{\mathbf{x}}_0), \tag{15}$$

where $\mathbf{x}_0$ denotes the principal point, $\tilde{\mathbf{x}}_a$, $\tilde{\mathbf{x}}_p$ and $\tilde{\mathbf{x}}_0$ represent row vector of dehomogenized coordinates, and Δ and d_0 can be understood as the depth relief and the average depth given the imaging scene. Suppose we are given a correspondence $(\mathbf{x}_p, \mathbf{x}'_p)$ from two general cameras, the corresponding linear embedding is $\mathbf{d} = [\tilde{\mathbf{x}}_p, \tilde{\mathbf{x}}'_p, 1]^T$. Based on the observation in (15), $\mathbf{d}$ can be understood from another point of view: $\mathbf{d} = [\tilde{\mathbf{x}}_a + \epsilon, \tilde{\mathbf{x}}'_a + \epsilon', 1]^T$, where $\epsilon = \frac{\Delta}{d_0}(\tilde{\mathbf{x}}_0 - \tilde{\mathbf{x}}_p)$ and

Algorithm 3. Homography Estimation with Linear Embedding

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

1: Apply Algorithm 2 on $\mathcal{S}$ to find an inlier set $\mathcal{I}$.
2: Post-processing on $\mathcal{I}$ to determine final estimation result $\mathbf{s}$.
3: **return** $\mathbf{s}$.

$\epsilon' = \frac{\Delta}{d_0}(\tilde{\mathbf{x}}'_0 - \tilde{\mathbf{x}}'_p)$ represent the noise proportional to $\frac{\Delta}{d_0}$. As $\mathbf{x}_a$ and $\mathbf{x}'_a$ are related by an affine model, if $\frac{\Delta}{d_0}$ is sufficiently small, we can still use Algorithm 2 with simple linear embedding of correspondences to exploit the structure, due to the noise-robust property of the ℓ_1 based objective. In this case, most inliers can be detected by Algorithm 2 if the camera is far from scene (d_0 is large). Otherwise, the detected inliers will generally lie closely to a plane that exhibits a small depth relief (Δ is small) and may only include a subset of the true inliers.

In a different point of view, we can also explain the results from the models themselves. Although affine model cannot describe exactly the perspective plane-plane transformation, it can be seen as the linear approximation of the non-linear homography model. This indicates that the affine model can be applied at least locally for the correspondences. In this case, the inliers that are spatially adjacent in one image can be detected by Algorithm 2.

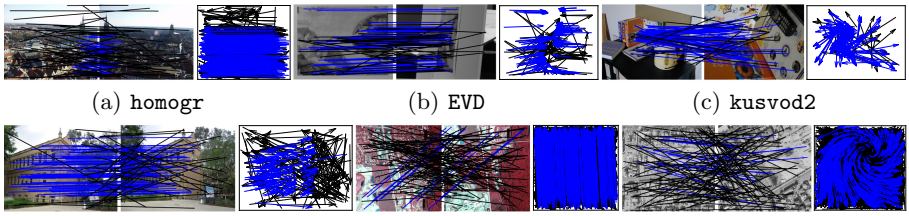
Concluding from the discussions above, we can see that in general scenarios, Algorithm 2 can be applied to detect at least a subset of the inliers. This is because the ℓ_1-based objective is more insensitive to the error induced by approximation. However, there is still a gap between the detected inliers and a model of good quality. In this paper, we show that the gap can be filled by applying local refinement technique to find the optimal model. This leads to the following extensions.

(1) *Homography Estimation.* As discussed above, when the scenes conform to a more general homography transformation, Algorithm 2 is not the ideal choice. However, we can expect that at least a subset of the inliers can be detected by Algorithm 2. In this case, we choose to subsequently run a local optimization step to include more inliers and recover the true homography. The detail is discussed in our post-processing procedure in Sect. 3.4. Some illustrative examples are given in Sect. 4.1. The algorithm is outlined in Algorithm 3.
(2) *Fundamental Matrix Estimation.* For fundamental matrix estimation task, it is generally unsolvable if we only have a group of affine or homography related correspondences [12]. To this end, we propose to detect two groups of inliers of disjoint planes, like a sequential RANSAC method [34]. In the first iteration, a group of inliers are detected by Algorithm 2, subsequently, the first group of inliers are excluded to detect the second group of inliers. Finally, fundamental matrix estimation can be achieved from the inliers of the two groups combined. Some illustrative examples are given in Sect. 4.1. The algorithm is outlined in Algorithm 4.

Algorithm 4. Fundamental Matrix Estimation with Linear Embedding

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

1: Apply Algorithm 2 on $\mathcal{S}$ to find an inlier set $\mathcal{I}_1$.
2: Exclude $\mathcal{I}_1$ from $\mathcal{S}$ to form $\mathcal{S}'$.
3: Apply Algorithm 2 on $\mathcal{S}'$ to find an inlier set $\mathcal{I}_2$.
4: Post-processing on $\mathcal{I}_1 \bigcup \mathcal{I}_2$ to determine final estimation result $\mathbf{s}$.
5: **return** $\mathbf{s}$.



(a) `homogr` (b) `EVD` (c) `kusvod2`

(d) `AdelaideRMF` (e) `RS`

Fig. 1. Some illustrative examples of our SRE. For the image pair, only a subset of 100 correspondences are shown for visibility. A motion field is additionally shown for each pair, where the head and tail of each arrow correspond to the positions of feature points in two images. The identified inliers are drawn in blue and outliers in black. (Color figure online)

3.4 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. After obtaining the inlier set $\mathcal{I}$, since the detected inliers may still include a number of outliers, and to achieve better accuracy, we run a fixed 100 samples from $\mathcal{I}$ and derive the models (homography, fundamental matrix) using DLT from them. Each model is evaluated on the original correspondence set and each so-far-the-best model is refined by a local optimization step [19]. Finally, we take the best model with the largest consensus as the output estimation result. Note that this strategy also require a predefined inlier-outlier threshold as the random sampling techniques do.

4 Experimental Results

In this section, we investigate the performance of the proposed method on real image data for the geometric estimation task. We name our method subspace recovery estimator (SRE). Some illustrative examples are shown in Fig. 1.

Fig. 2. Illustrative examples from `homogr` of SRE for homography estimation. The first row represents the detected inliers by Algorithm 2, and the second row represents the inliers after post-processing to find a homography.

Homography Estimation. The widely used `homogr`[1] and the `EVD`[2] dataset are adopted. The `homogr` dataset contains 16 image pairs of relatively short baselines, while the `EVD` dataset contains 15 image pairs undergoing extremely view changes, *i.e.*, wide baselines. Both datasets are provided with a number of manually selected true correspondences for model evaluation. Additionally, we also collect 20 pairs of remote sensing images to create the `RS` dataset. Since the imaging equipment is very far from the scene in remote sensing, it conforms to the affine camera model almost perfectly. The `RS` dataset is featured by a high outlier ratio (above 80% in average), which serves to test the robustness of each method under extreme outliers. The inliers are manually labeled for this dataset.

Fundamental Matrix Estimation. The widely used `kusvod2`[3] and the `AdelaideRMF`[4] dataset are adopted. The `kusvod2` contains 16 image pairs of both weak and strong perspectives, and a number of true correspondences are provided for model evaluation. 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 of it undergoing only a single motion. The image pairs are generally of weak perspective since the camera is distant to the building.

4.1 Qualitative Analysis of Linear Embedding

Our SRE involves several strategies for geometric estimation, based on the exploition of different embeddings of correspondences, *i.e.* (4), (5), and (11). While (4) and (5) is well-grounded since they are derivatives of the classic DLT, the efficacy of linear embedding (11) requires to be further investigated for general scenes.

The illustrative examples for homography estimation with linear embedding, *i.e.* Algorithm 3, are shown in Fig. 2. It can be seen that the detected inliers may only be a subset of the true inliers, and even include some outliers. However, after the post-processing involving local optimization steps, all inliers can be

[1] http://cmp.felk.cvut.cz/data/geometry2view/.
[2] http://cmp.felk.cvut.cz/wbs/.
[3] http://cmp.felk.cvut.cz/data/geometry2view/.
[4] https://cs.adelaide.edu.au/~hwong/doku.php?id=data.

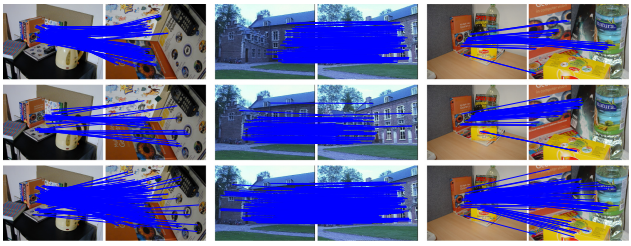


Fig. 3. Illustrative examples from `kusvod2` of SRE for fundamental matrix estimation. The first two rows represent the first and second groups of detected inliers by Algorithm 2. The last row represents the inliers after post-processing to find a fundamental matrix.

found and the true homography can be then recovered. The illustrative examples for fundamental matrix estimation with linear embedding, *i.e.* Algorithm 4, are shown in Fig. 3. Clearly, for a single run of Algorithm 2, the detected inliers generally lie closely to a plane. However, by running Algorithm 2 iteratively to find the second group of inliers and merging the two groups, the inlier set is then sufficiently diverse, leading to accurate estimation results after post-processing. Note that the detected inliers do not necessarily lie closely to a natural plane of the 3D object, practically, they may also be grouped by a virtual plane in the 3D space.

4.2 Fundamental and Homography Estimation

In our SRE, embeddings based on fundamental matrix and homography estimation in DLT, *i.e.* (4) and (5), can be used as in Algorithm 1. The linear embedding, *i.e.* (11), can be utilized in Algorithm 2. Another embedding of interest is (10), which can be used in a similar way to Algorithm 1. We denote these variants as SRE-F, SRE-H, SRE-A and SRE-At, respectively. For SRE-F, the detected inliers are directly used for homography and fundamental matrix estimation. For SRE-H, SRE-A and SRE-At, a single run is conducted for homography estimation, and two runs to find two groups of inliers are conducted for fundamental matrix estimation.

We compare our SRE with the baseline RANSAC [15], and the state-of-the-art methods USAC [29] and MAGSAC [2]. The parameters are setting according to the original papers. The number of maximum trials is set to $5,000$ for RANSAC, MAGSAC and USAC1, and $50,000$ for USAC2. For our SRE, we empirically set the threshold as $1/5$ of the mean residual to identify inliers. The inlier-outlier threshold is set to 2 pixels for all methods.

We use the average geometric error of the given inliers *w.r.t.* the estimated model as the evaluation metric. As a failed model would induce unreliable statistics, we exclude the failed cases to compute the average error ($\mathbf{e}$) and also report the proportion of failures ($\mathbf{f}$). The geometric error is computed as the Sampson distance. To determine a failed model, we use two thresholds, *i.e.*, 5 pixels ($\mathbf{e}_1$,

Table 1. Quantitative evaluation. The datasets, problems, numbers of image pairs and metrics are shown in the first three columns. The other columns show the average mean geometric error (**e**) and proportion of failures (**f**) over 100 runs, where the subscripts 1 and 2 denote the results given 5 and 10 pixels as the threshold to determine a failed model, respectively. The mean processing time (in milliseconds), *i.e.* **t**, and the summary statistics of all datasets, *i.e.* `all`, are given. Bold indicates the best result.

Alg.			RANSAC	USAC1	USAC2	MAGSAC	SRE-F	SRE-H	SRE-At	SRE-A
`homogr`	**H**,16	$\mathbf{e}_1$	1.73	1.41	1.40	1.69	1.14	1.15	**1.12**	1.15
		$\mathbf{f}_1$	0.055	0.141	0.100	0.210	0.078	0.163	0.148	**0.036**
		$\mathbf{e}_2$	1.92	1.56	1.53	1.87	**1.31**	1.37	1.34	**1.31**
		$\mathbf{f}_2$	0.016	0.116	0.077	0.181	0.046	0.124	0.108	**0.001**
		t	207.1	**20.9**	37.3	185.2	69.0	70.6	66.2	68.4
`EVD`	**H**,15	$\mathbf{e}_1$	1.91	1.04	1.02	1.31	1.04	1.00	**0.93**	1.02
		$\mathbf{f}_1$	0.255	0.734	0.733	**0.208**	0.263	0.367	0.506	0.216
		$\mathbf{e}_2$	2.30	1.06	**1.02**	1.50	1.08	1.34	1.39	1.08
		$\mathbf{f}_2$	0.189	0.733	0.733	**0.181**	0.259	0.334	0.471	0.208
		t	343.9	**29.3**	196.9	213.8	54.1	58.5	58.3	50.4
`kusvod2`	**F**,16	$\mathbf{e}_1$	1.65	1.56	1.52	1.03	0.81	0.80	**0.76**	0.81
		$\mathbf{f}_1$	0.143	0.094	0.085	0.151	0.139	0.132	0.137	**0.084**
		$\mathbf{e}_2$	1.94	1.77	1.70	1.32	1.13	0.95	**0.85**	0.87
		$\mathbf{f}_2$	0.096	0.056	**0.053**	0.109	0.094	0.112	0.125	0.076
		t	13.8	**16.4**	16.8	338.9	32.2	36.7	34.6	38.9
`Adelaide`	**F**,19	$\mathbf{e}_1$	0.76	0.65	0.63	0.89	0.63	0.62	0.58	**0.55**
		$\mathbf{f}_1$	0.002	0.001	**0.000**	0.084	0.068	0.117	0.094	**0.000**
		$\mathbf{e}_2$	0.78	0.66	0.63	1.17	0.83	0.71	0.66	**0.55**
		$\mathbf{f}_2$	**0.000**	**0.000**	**0.000**	0.041	0.040	0.106	0.084	**0.000**
		t	41.8	**26.2**	37.8	290.5	33.3	35.3	30.4	33.0
`RS`	**H**,20	$\mathbf{e}_1$	1.65	**0.94**	1.15	1.95	1.68	2.11	1.96	2.22
		$\mathbf{f}_1$	0.349	0.603	0.227	0.763	0.618	0.818	0.805	**0.085**
		$\mathbf{e}_2$	2.17	**1.00**	1.18	2.90	1.96	2.26	2.23	1.59
		$\mathbf{f}_2$	0.182	0.599	0.221	0.709	0.590	0.812	0.791	**0.000**
		t	945.4	**23.6**	92.7	862.0	136.2	151.1	145.2	117.4
`all`		$\mathbf{e}_1$	1.48	1.12	1.13	1.25	0.98	0.94	**0.90**	0.94
		$\mathbf{f}_1$	0.163	0.312	0.215	0.300	0.245	0.335	0.349	**0.079**
		$\mathbf{e}_2$	1.87	1.22	1.21	1.57	1.18	1.13	1.23	**1.09**
		$\mathbf{f}_2$	0.096	0.299	0.203	0.259	0.217	0.314	0.328	**0.051**
		t	330.2	**23.4**	74.2	399.5	67.2	72.9	69.3	62.5

$\mathbf{f}_1$) and 10 pixels ($\mathbf{e}_2$, $\mathbf{f}_2$). If the estimated model induces an average geometric error larger than the threshold, it is then deemed as a failed one. For each method, we report the statistics from 100 runs on each image pair.

The evaluation results are given in Table 1. It can be seen that RANSAC is quite robust compared to USAC and MAGSAC, resulting in a low propor-

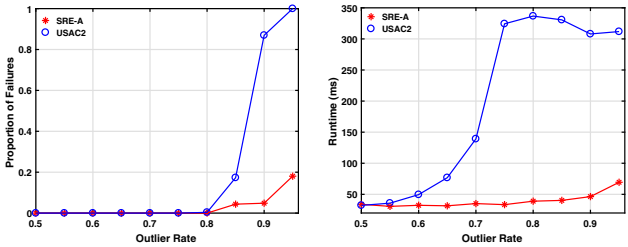


Fig. 4. The performance comparison of our SRE-A and USAC2 on synthetic data. The left figure gives the result in terms of proportion of failures, and the right gives the result in terms of proportion of runtime.

tion of failures. However, it is generally less accurate and very time consuming for geometric estimation task. USAC unifies many important advancements of the random sampling based estimators, and achieves significant improvements in terms of both accuracy and efficiency. However, the robustness seems to be impaired, and its performance on the challenging datasets, *i.e.* `EVD` and `RS` is very limited. MAGSAC was proposed to get rid of the need to specify an inlier-outlier threshold. It performs well on the `EVD` dataset, however, its robustness is not consistent on the other datasets, and cannot handle the challenging `RS`. For all the four variants of our SRE, the efficiency is advantaged and comparable to USAC. Specifically, SRE-F and SRE-H are often more accurate, but the robustness is not advantaged. SRE-A outperforms SRE-At by a large margin, since the induced subspace recovery problem is much easier to solve. Overall, the most effective variant of our SRE is SRE-A. It is the most robust in average, and comparatively efficient to USAC with preferable accuracy.

To conclude, SRE-A is the best performer among all the variants of our SRE. Compared to the current state-of-the-art geometric estimation algorithms, it is the most robust one with preferred accuracy. The efficiency is comparable to the fastest USAC. A nice property of SRE-A is its robustness to outliers, as demonstrated on `RS` dataset where it shows significantly better results.

4.3 Sensitivity to Outlier Rate

Since the goal of our algorithm is to conduct geometric estimation in the presence of outliers, a straightforward question is its sensitivity to outlier rate. For investigation, we use the `AdelaideRMF` dataset since the inlier correspondences are annotated. We generate the outlier correspondences by matching two random points in the two images and control the outlier rate by adding a number of randomly generated outliers to the inliers. In the experiment, 100 runs are conducted for each algorithm on each instance to give the average performance. We use the best performer SRE-A, and the USAC algorithm with maximum trials of $50,000$ (USAC2) for comparison. The proportion of failures (with 5 pixels as threshold) and runtime are used for evaluation, and the results are presented in Fig. 4.

From the results in Fig. 4, we can see that our SRE-A outperforms USAC in both robustness (smaller proportion of failures) and efficiency. Our method can still work under extreme outliers (up to 95%), whereas USAC2 tends to fail in such circumstances. Also, the efficiency of our method is unaffected by the outlier rate. This is a significant advantage since the runtime of USAC grows exponentially with the outlier rate increasing, until the maximum trials is approached.

5 Conclusion

In this paper, we propose a novel method SRE for geometric estimation. With a robust ℓ_1-based objective, the intrinsic linear structure is explored, and several efficient algorithms are designed for geometric estimation with robust subspace recovery technique. Experiments on real-world image data demonstrate the superiority of the proposed SRE, in terms of both robustness and accuracy compared to the state-of-the-art.

References

1. Barath, D., Matas, J.: Graph-cut RANSAC. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6733–6741 (2018)
2. Barath, D., Matas, J., Noskova, J.: MAGSAC: marginalizing sample consensus. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 10197–10205 (2019)
3. Bazin, J.C., Li, H., Kweon, I.S., Demonceaux, C., Vasseur, P., Ikeuchi, K.: A branch-and-bound approach to correspondence and grouping problems. IEEE Trans. Pattern Anal. Mach. Intell. **35**(7), 1565–1576 (2012)
4. Bazin, J.-C., Seo, Y., Hartley, R., Pollefeys, M.: Globally optimal inlier set maximization with unknown rotation and focal length. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8690, pp. 803–817. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10605-2_52
5. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. Int. J. Comput. Vis. **74**(1), 59–73 (2007)
6. Cai, Z., Chin, T.J., Koltun, V.: Consensus maximization tree search revisited. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1637–1645 (2019)
7. Cai, Z., Chin, T.J., Le, H., Suter, D.: Deterministic consensus maximization with biconvex programming. In: Proceedings of the European Conference on Computer Vision, pp. 685–700 (2018)
8. Campbell, D., Petersson, L., Kneip, L., Li, H.: Globally-optimal inlier set maximisation for simultaneous camera pose and feature correspondence. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1–10 (2017)
9. Chum, O., Matas, J.: Matching with PROSAC-progressive sample consensus. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 220–226 (2005)
10. Chum, O., Matas, J.: Optimal randomized RANSAC. IEEE Trans. Pattern Anal. Mach. Intell. **30**(8), 1472–1482 (2008)

11. Chum, O., Matas, J., Kittler, J.: Locally optimized RANSAC. In: Michaelis, B., Krell, G. (eds.) DAGM 2003. LNCS, vol. 2781, pp. 236–243. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45243-0_31
12. Chum, O., Werner, T., Matas, J.: Two-view geometry estimation unaffected by a dominant plane. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 772–779 (2005)
13. Ding, T., et al.: Noisy dual principal component pursuit. In: Proceedings of the International Conference on Machine Learning, pp. 1617–1625 (2019)
14. Enqvist, O., Ask, E., Kahl, F., Åström, K.: Tractable algorithms for robust model estimation. Int. J. Comput. Vis. **112**(1), 115–129 (2015)
15. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
16. Fragoso, V., Sen, P., Rodriguez, S., Turk, M.: EVSAC: accelerating hypotheses generation by modeling matching scores with extreme value theory. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2472–2479 (2013)
17. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2003)
18. Le, H.M., Chin, T.J., Eriksson, A., Do, T.T., Suter, D.: Deterministic approximate methods for maximum consensus robust fitting. IEEE Trans. Pattern Anal. Mach. Intell. (2019)
19. Lebeda, K., Matas, J., Chum, O.: Fixing the locally optimized RANSAC-full experimental evaluation. In: Proceedings of the British Machine Vision Conference, pp. 1–11 (2012)
20. Lerman, G., Maunu, T.: An overview of robust subspace recovery. Proc. IEEE **106**(8), 1380–1410 (2018)
21. Lerman, G., McCoy, M.B., Tropp, J.A., Zhang, T.: Robust computation of linear models by convex relaxation. Found. Comput. Math. **15**(2), 363–410 (2015)
22. Li, H.: Consensus set maximization with guaranteed global optimality for robust geometry estimation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1074–1080 (2009)
23. Ma, J., Zhao, J., Jiang, J., Zhou, H., Guo, X.: Locality preserving matching. Int. J. Comput. Vis. **127**(5), 512–531 (2019)
24. Ma, J., Zhao, J., Tian, J., Yuille, A.L., Tu, Z.: Robust point matching via vector field consensus. IEEE Trans. Image Process. **23**(4), 1706–1721 (2014)
25. Matas, J., Chum, O.: Randomized RANSAC with Td, d test. Image Vis. Comput. **22**(10), 837–842 (2004)
26. Mur-Artal, R., Montiel, J.M.M., Tardos, J.D.: ORB-SLAM: a versatile and accurate monocular slam system. IEEE Trans. Rob. **31**(5), 1147–1163 (2015)
27. Ni, K., Jin, H., Dellaert, F.: GroupSAC: efficient consensus in the presence of groupings. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2193–2200 (2009)
28. Purkait, P., Zach, C., Eriksson, A.: Maximum consensus parameter estimation by reweighted ℓ_1 methods. In: Pelillo, M., Hancock, E. (eds.) EMMCVPR 2017. LNCS, vol. 10746, pp. 312–327. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-78199-0_21
29. Raguram, R., Chum, O., Pollefeys, M., Matas, J., Frahm, J.M.: USAC: a universal framework for random sample consensus. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 2022–2038 (2012)
30. Torr, P.H.S.: Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. Int. J. Comput. Vis. **50**(1), 35–61 (2002)

31. Torr, P.H., Nasuto, S.J., Bishop, J.M.: NAPSAC: high noise, high dimensional robust estimation-it's in the bag. In: British Machine Vision Conference (BMVC) (2002)
32. Torr, P.H., Zisserman, A.: MLESAC: a new robust estimator with application to estimating image geometry. Comput. Vis. Image Underst. **78**(1), 138–156 (2000)
33. Tsakiris, M.C., Vidal, R.: Dual principal component pursuit. J. Mach. Learn. Res. **19**(1), 684–732 (2018)
34. Vincent, E., Laganiére, R.: Detecting planar homographies in an image pair. In: Proceedings of the International Symposium on Image and Signal Processing and Analysis, pp. 182–187 (2001)
35. Zhu, Z., Wang, Y., Robinson, D., Naiman, D., Vidal, R., Tsakiris, M.: Dual principal component pursuit: improved analysis and efficient algorithms. In: Advances in Neural Information Processing Systems, pp. 2171–2181 (2018)