Efficient Algorithms for Global Outlier Removal in Large-scale Structure-from-Motion

Fei Wen, Danping Zou, Ruihang Miao, Rendong Ying, and Peilin Liu

Department of Electronic Engineering

Shanghai Jiao Tong University

Shanghai 200240, China

{*wenfei, dpzou, rdying, liupeilin*}@*sjtu.edu.cn*

Abstract—This work addresses the outlier removal problem in large-scale global structure-from-motion. In such applications, outlier removal is very useful to mitigate the deterioration caused by mismatches in the feature point matching step. Unlike existing outlier removal methods, we exploit the structure in multiview geometry problems to propose a dimension reduced formulation, based on which two efficient methods have been developed. The first method considers a convex relaxed ℓ_1 minimization and is solved by a single linear programming (LP). The second method approximately solves the ideal ℓ_0 minimization by an iteratively reweighted method. The dimension reduction results in a significant speedup of the new algorithms. Further, the iteratively reweighted method can significantly reduce the possibility of removing true inliers. Results show that, compared with state-of-the-art algorithms (e.g., the ℓ_1 method), the proposed algorithms are more than three times faster and meanwhile produce better consensus sets. Matlab code for reproducing the results is available at *https://github.com/FWen/OUTLR.git*.

Index Terms—Structure from motion, multiview reconstruction, large scale, robust fitting, outlier.

I. INTRODUCTION

Structure-from-Motion (SfM) tries to recover the 3D point clouds and camera poses from a set of unordered images. There are two approaches to this problem. The first one is the incremental approach. It starts from two images, and grows the point clouds and computes the poses incrementally. The second one is the global approach, which firstly estimates the relative orientations among different images and then recovers both the 3D point positions and camera positions simultaneously. The recovered structures from both approaches are used as initialization in bundle adjustment to get the final results. It has been shown in [28] that for large-scale SfM problems, the global approach works significantly better, since the global approach formulates SfM as a convex optimization problem, which guarantees the global optimum result theoretically.

Though outlier removal has been applied to remove the incorrect point correspondences in the first step of global SfM, which solves the relative orientations from pairwise matching between different views, a large number of outliers still exist because the epipolar constraints among local views usually do not reveal outliers in long point tracks. Those outliers will significantly degrade the structure estimation if they are not well processed [19]. The main interest of this work is to find an efficient and robust solution to address the outlier removal problem in global SfM.

The problem of outlier removal in model fitting, or robust model fitting, is fundamentally critical to many computer vision applications, including fundamental matrix estimation, homography matrix estimation, vision-based robotics navigation, and global outlier removal in SfM [1]–[3], [29]. Given a set of M measurements $\{\mathbf{a}_i, y_i\}_{i=1,\cdots M}$, an important problem arises in many computer vision applications is to remove the outliers in the data set. It is also known as the maximum consensus problem [4], which aims to find a model, parameterized by $\mathbf{x} \in \mathbb{R}^N$, that is consistent with as many of the input data as possible, i.e., has the largest consensus set I as [5]

$$
\begin{aligned}
&\underset{\mathbf{x}, I \subseteq \Omega}{\text{maximize}} \, |I| \\
&\text{subject to} \quad \left| \mathbf{a}_i^T \mathbf{x} - y_i \right| \leq \delta, \quad \forall i \in I
\end{aligned}
\tag{1}
$$

where $\delta > 0$ is the inlier threshold, $\Omega = \{1, 2, \cdots, M\}$ is the index set. For a solution I^* with size $|I^*|$, I^* denotes the index set of the true inliers, and $\Omega \backslash I^*$ denotes the index set of the true outliers.

On the one hand, due to the intractability of the robust geometric fitting problem, the global optimum can only be found by searching [25], which makes globally optimal algorithms only suitable for low-dimensional problems. On the other hand, the class of randomized hypothesize-and-verify algorithms are more efficient and popular, e.g., RANSAC [26] and its many variants [27]. Although such randomized algorithms are efficient, they can usually produce only approximate solution and do not guarantee a good estimate due to their randomized nature. Very recently, deterministic methods fill the gap between these two classes have been proposed in [2], [4], [24]. Such deterministic methods are much more efficient than globally optimal algorithms, while being able to achieve better solution quality compared with hypothesize-and-verify based randomized algorithms.

Due to the high-dimensionality of the global outlier removal problem in global SfM, the methods in [4], [5], [24]–[27] are usually impractical and not applicable. The most efficient algorithm suitable for this problem is the ℓ_1 method [6], which solves a convex relaxation of (1). Although there is no theoretical guarantee of success, the ℓ_1 method works very well in practice and can yield a significant drop in reprojection error.

In this work, we propose outlier removal algorithms that are

more efficient than the ℓ_1 method [6] and suitable for large-scale Global SfM. The key idea is that we use a dimension reduced formulation to reduce the computational complexity. Similar to [6], we present two versions of outlier removal algorithms. The first fast version formulates the outlier removal as a convex relaxed ℓ_1 minimization that is solved by linear programming (LP). The second one approximately solves the ideal ℓ_0 formulation by an iteratively reweighted algorithm with only a little bit of efficiency trade-off.

We have conducted experiments on both synthetic and real world datasets. Experimental results show that our methods retain more inliers than existing methods, and run much faster.

II. PROPOSED FORMULATION AND CONNECTION TO EXISTING WORKS

A. Proposed Formulation

Using a nonnegative auxiliary variable $\mathbf{s} \in \mathbb{R}^M$, $\mathbf{s} \geq 0$, problem (1) can be recast into the following formulation

$$\begin{aligned} \underset{\mathbf{x}, \mathbf{s}}{\text{maximize}} \ & M - \|\mathbf{s}\|_0 \\ \text{subject to} \ & \left|\mathbf{a}_i^T \mathbf{x} - y_i\right| \leq \delta + s_i, \quad s_i \geq 0 \end{aligned} \tag{2}$$

where $\|\mathbf{s}\|_0$ denotes the ℓ_0 norm which counts the number of nonzero elements in $\mathbf{s}$. For a solution $\mathbf{s}^*$ of (2), it holds $\text{supp}(\mathbf{s}^*) = \Omega \backslash I^*$, which is the index set of the outliers, and $M - \|\mathbf{s}\|_0 = |I^*|$. $\text{supp}(\cdot)$ denotes the support set (the index set of nonzero) of a vector. Equivalently, problem (2) can be expressed as the constrained ℓ_0 minimization

$$\begin{aligned} \underset{\mathbf{x}, \mathbf{s}}{\text{minimize}} \ & \|\mathbf{s}\|_0 \\ \text{subject to} \ & \left|\mathbf{a}_i^T \mathbf{x} - y_i\right| \leq \delta + s_i, \quad s_i \geq 0. \end{aligned} \tag{3}$$

The constraint in (3) can be expressed as linear inequalities. Specifically, the inequality constraint in (3) is equivalent to

$$\mathbf{a}_i^T \mathbf{x} - y_i \leq \delta + s_i \quad \text{and} \quad -\mathbf{a}_i^T \mathbf{x} + y_i \leq \delta + s_i. \tag{4}$$

Let $\mathbf{A} = \left[\mathbf{A}_1^T, \mathbf{A}_2^T, \cdots, \mathbf{A}_M^T\right]^T$ and $\mathbf{b} = \left[\mathbf{b}_1^T, \mathbf{b}_2^T, \cdots, \mathbf{b}_M^T\right]^T$ with $\mathbf{A}_i = [\mathbf{a}_i, -\mathbf{a}_i]^T$ and $\mathbf{b}_i = [y_i + \delta, -y_i + \delta]^T$. Then, the formulation (3) can be rewritten as

$$\begin{aligned} \underset{\mathbf{x}, \mathbf{s}}{\text{minimize}} \ & \|\mathbf{s}\|_0 \\ \text{subject to} \ & \mathbf{A}\mathbf{x} \leq \mathbf{b} + \mathbf{s} \otimes \mathbf{1}_{\kappa \times 1}, \quad \mathbf{s} \geq 0 \end{aligned} \tag{5}$$

where $\kappa = 2$, $\mathbf{1}_{\kappa \times 1}$ is a κ-dimensional vector with all elements be 1, and $\otimes$ denotes the Kronecker product.

B. Multiview Geometry

The above formulations are designed for the linear regression residual $\left|\mathbf{a}^T \mathbf{x} - y\right|$, which can be extended to handle geometric residuals, e.g., in multiview reconstruction. In multiview geometry, the goal is to estimate the structure of the scene and the camera motion from image projections. For instance, let (z_1^i, z_2^i) be a measurement in one of the images and $\mathbf{z}$ be its corresponding 3D-point. Given the camera matrix $\mathbf{P} = [\mathbf{R}, \mathbf{t}] \in \mathbb{R}^{3 \times 4}$, the squared reprojection error is [6]

$$E_i(\mathbf{z}, \mathbf{P}) = \left\|z_1^i - \frac{\mathbf{r}_1 \mathbf{z} + t_1}{\mathbf{r}_3 \mathbf{z} + t_3}, z_2^i - \frac{\mathbf{r}_2 \mathbf{z} + t_2}{\mathbf{r}_3 \mathbf{z} + t_3}\right\|_2^2 \tag{6}$$

where $\mathbf{r}_j$ denotes the j-th row of $\mathbf{R}$, $\mathbf{t} = [t_1, t_2, t_3]^T \in \mathbb{R}^3$ denotes the translation of the camera. In the general framework, a reformulation of the squared error residuals is of the form

$$E_i(\mathbf{x}) = \frac{\left\|\mathbf{u}_i^T \mathbf{x} + \tilde{u}_i, \mathbf{v}_i^T \mathbf{x} + \tilde{v}_i\right\|_2^2}{\left(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i\right)^2}. \tag{7}$$

In many geometry problems in computer vision, if either $\mathbf{z}$ or $\mathbf{R}$ is known, the residual formulation (6) can be rewritten in the form of (7). For example, for the triangulation problem, $\mathbf{R}$ and $\mathbf{t}$ are kept fixed, and $\mathbf{x}$ denotes the position parameters of 3D-points. For the multiview reconstruction problem, both the positions of the 3D-points and the positions of the cameras are unknown. In this case, $\mathbf{x}$ in (7) contains the parameters of the 3D-points and the camera translations, which is also called the known rotation problem [7].

To handle such geometric residuals, in addition to the Euclidian distance, the coordinate-wise max distance (ℓ_∞ norm) and the absolute distance (ℓ_1 norm) are also popular due to their convenience. Specifically, a general formulation of the error residuals is of the form [7], [8]

$$E_i(\mathbf{x}) = \frac{\left\|\mathbf{u}_i^T \mathbf{x} + \tilde{u}_i, \mathbf{v}_i^T \mathbf{x} + \tilde{v}_i\right\|_p}{\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i} \quad \text{with} \quad \mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i > 0$$

where $E_i(\mathbf{x})$ is quasi-convex for any $p \geq 1$ [16]. With an inlier threshold $\delta > 0$, the constraint for outlier removal is

$$\left\|\mathbf{u}_i^T \mathbf{x} + \tilde{u}_i, \mathbf{v}_i^T \mathbf{x} + \tilde{v}_i\right\|_p \leq \delta(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i) \tag{8}$$

where $\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i > 0$ is naturally satisfied as $\delta > 0$.

For the ℓ_1 norm, i.e., $p = 1$, the constraint becomes

$$|\mathbf{u}_i^T \mathbf{x} + \tilde{u}_i| + |\mathbf{v}_i^T \mathbf{x} + \tilde{v}_i| \leq \delta(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i)$$

which can be expressed as the following linear inequalities

$$\begin{aligned} (\mathbf{u}_i^T + \mathbf{v}_i^T)\mathbf{x} + \tilde{u}_i + \tilde{v}_i &\leq \delta(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i) \\ (\mathbf{u}_i^T - \mathbf{v}_i^T)\mathbf{x} + \tilde{u}_i - \tilde{v}_i &\leq \delta(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i) \\ (\mathbf{v}_i^T - \mathbf{u}_i^T)\mathbf{x} - \tilde{u}_i + \tilde{v}_i &\leq \delta(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i) \\ -(\mathbf{u}_i^T + \mathbf{v}_i^T)\mathbf{x} - \tilde{u}_i - \tilde{v}_i &\leq \delta(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i). \end{aligned}$$

Meanwhile, in practical applications, a depth constraint can be additionally considered as

$$d_{\min} \leq \mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i \leq d_{\max}$$

where $d_{\min}$ and $d_{\max}$ denote the minimal and maximal depth, respectively. It can be converted into two linear constraints

$$-\mathbf{w}_i^T \mathbf{x} - \tilde{\omega}_i \leq -d_{\min} \quad \text{and} \quad \mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i \leq d_{\max}$$

In this case, $\kappa = 6$ in the constraint of (5).

For the ℓ_∞ norm, i.e., $p = \infty$, the constraint (8) becomes

$$\max\left\{|\mathbf{u}_i^T \mathbf{x} + \tilde{u}_i|, |\mathbf{v}_i^T \mathbf{x} + \tilde{v}_i|\right\} \leq \varepsilon(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i)$$

which is equivalent to

$$|\mathbf{u}_i^T \mathbf{x} + \tilde{u}_i| \leq \delta(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i) \quad \text{and} \quad |\mathbf{v}_i^T \mathbf{x} + \tilde{v}_i| \leq \delta(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i).$$

Using (4), equivalent linear inequalities of the constraint (8)

can be derived as follows

$$\mathbf{u}_i^T \mathbf{x} + \tilde{u}_i \le \delta(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i), \quad -\mathbf{u}_i^T \mathbf{x} - \tilde{u}_i \le \delta(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i)$$

$$\mathbf{v}_i^T \mathbf{x} + \tilde{v}_i \le \delta(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i), \quad -\mathbf{v}_i^T \mathbf{x} - \tilde{v}_i \le \delta(\mathbf{w}_i^T \mathbf{x} + \tilde{\omega}_i).$$

The effectiveness of the ℓ_∞ norm has been demonstrated in various problems in computer vision [9]–[11].

C. Connection to Existing Works

A reformulation of (1) has been considered in [5] as

$$\begin{aligned} \underset{\mathbf{x},\mathbf{z}}{\text{minimize}} \quad & \sum_i z_i \\ \text{subject to} \quad & \left| \mathbf{a}_i^T \mathbf{x} - y_i \right| \le \delta + z_i L, \quad z_i \in \{0,1\} \end{aligned} \quad (9)$$

where L is a large positive constant. For a solution $\mathbf{z}^*$ of (9), it is easy to see that $\text{supp}(\mathbf{z}^*) = \text{supp}(\mathbf{s}^*)$, which implies the equivalence of the formulations (2), (3), (5) and (9) for solving the maximum consensus problem (1). To solve the maximum consensus problem exactly and efficiently, a guaranteed outlier removal (GORE) approach based on mixed integer linear programming has been proposed in [5] to reduce the runtime of exact algorithms. But it does not scale to high-dimensional problems, e.g., large-scale multiview reconstruction.

Very recently, deterministic approximate methods have been proposed in [4], [24]. These methods reformulate the consensus maximization problem with linear complementarity constraints, and employ the Frank-Wolfe optimization scheme and alternating direction method of multipliers (ADMM) to efficiently solve the reformulations. These algorithms are efficient and effective for low-dimensional problems such as fundamental matrix and homography estimation, but they still do not scale to large-scale multiview reconstruction problems.

It is popular to solve convex relaxed formulations of (1), e.g., ℓ_1 approximation [6], [14], [15]. The most efficient ℓ_1 method [6] considers a formulation as

$$\begin{aligned} \underset{\mathbf{x},\mathbf{s}}{\text{minimize}} \quad & \mathbf{1}^T \tilde{\mathbf{s}} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} + \tilde{\mathbf{s}}, \quad \tilde{\mathbf{s}} \ge 0. \end{aligned} \quad (10)$$

This ℓ_1 minimization formulation can be viewed as a convex relaxation of (5), where the nonconvex ℓ_0 norm is replaced by its convex envelope, i.e., the ℓ_1 norm. Meanwhile, the structure in $\mathbf{s} \otimes \mathbf{1}_{\kappa \times 1}$ in the constraint is ignored and, hence, the number of unknown parameters in $\tilde{\mathbf{s}}$ is κM while that in $\mathbf{s}$ is M. Since in our formulation the dimension of variables is significantly reduced, it can be solved more efficiently with lower computational complexity (see section IV). The dimension reduction does not only reduce the computational complexity, but also yields improved results.

The ℓ_1 relaxation is convenient due to its convexity and that well-developed LP solvers can be directly applied. However, the convex relaxation may degrade the performance. It has been demonstrated in the sparse recovery researches that, the ℓ_0 or ℓ_q $(0 < q < 1)$ norm can usually yield a sparser solution than the ℓ_1 norm [12]. Empirical results have shown that the ℓ_1 minimization (10) is likely to remove inliers in some conditions [6], [24]. Since the ℓ_0 and ℓ_q norm penalties

tend to yield a sparser solution, they can be expected to reduce the possibility of removing true inliers. In this regard, a re-weighted ℓ_1 method has been proposed recently in [30].

III. PROPOSED ALGORITHMS

Generally, it is difficult to directly solve the nonconvex ℓ_0 minimization problem (5). In this section, we first propose an algorithm to solve a convex relaxed version of it. Then, we develop an iteratively reweighted algorithm to approximately solve the ideal ℓ_0 minimization problem (5).

A. ℓ_1 Algorithm with Reduced Dimension

We consider a convex relaxation of (5) via replacing the ℓ_0-norm by its convex envelope, the ℓ_1-norm, as

$$\begin{aligned} \underset{\mathbf{x},\mathbf{s}}{\text{minimize}} \quad & \|\mathbf{s}\|_1 \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} + \mathbf{s} \otimes \mathbf{1}_{\kappa \times 1}, \quad \mathbf{s} \ge 0. \end{aligned} \quad (11)$$

This formulation is similar to that considered in [6], except for that the number of slack variables in (11) is M while that in the ℓ_1 algorithm [6] is κM. The dimension reduction would result in a speedup. The dual problem of (11) is given by

$$\begin{aligned} \underset{\mathbf{y},\mathbf{v}}{\max} \quad & -\mathbf{b}^T \mathbf{y} \\ \text{subject to} \quad & \mathbf{A}^T \mathbf{y} = \mathbf{0} \\ & \mathbf{1}_{M \times 1} - \mathbf{J}\mathbf{y} - \mathbf{v} = \mathbf{0} \\ & \mathbf{y} \ge 0, \quad \mathbf{v} \ge 0 \end{aligned} \quad (12)$$

where $\mathbf{y} \in \mathbb{R}^{\kappa M}$ and $\mathbf{v} \in \mathbb{R}^M$ are the dual variables, and $\mathbf{J} = \mathbf{I}_M \otimes \mathbf{1}_{1 \times \kappa} \in \mathbb{R}^{M \times (\kappa M)}$. $\mathbf{I}_M$ is an identity matrix of size M. The dual problem (12) can be solved by well-developed LP solvers. The algorithm is summarized as follows.

Algorithm 1: ℓ_1 algorithm with reduced dimension

Input: The set of measurements Ω, inlier threshold $\delta > 0$.
Begin:
    Construct $\mathbf{A}$ and $\mathbf{b}$ from the measurements Ω.
    Solve the LP problem (12) to obtain $\mathbf{s}$.
    Remove the residuals for which $s_i > 0$.
End
Output: A subset $I \subseteq \Omega$ of the measurements for which
    $\exists \mathbf{x}$ such that $E_i(\mathbf{x}) \le \delta, \forall i \in I$.

B. Iteratively Reweighted Algorithm

As shown in [6], [24], the ℓ_1 method probably removes true inliers in practical applications. This is explained in the last section, as the ℓ_1 minimization may yield a solution not sparse enough. Ideally, we should minimize the number of nonzero elements in $\mathbf{s}$, i.e., the ℓ_0 minimization problem (5). However, exact solving of (5) is difficult. Inspired by the success of iteratively reweighted methods in sparse recovery researches [13], we propose an iteratively reweighted algorithm to approximately solve the intractable problem (5).

First, we approximate the ℓ_0 norm by the ℓ_q norm with a small value of q (e.g., $q = 0.1$ in the experiments). The ℓ_q

norm usually yields a sparser solution than ℓ_1 norm [20], [21]. Then, at the $(k+1)$-th iteration, the ℓ_q norm is approximated via first-order expansion (linearization) at $\mathbf{s}^k$ obtained at the k-th iteration as

$$\|\mathbf{s}\|_{q,\varepsilon}^q \approx \sum_{i=1}^M \left(\left|s_i^k\right| + \varepsilon\right)^{q-1} |s_i|$$

where ε is a small positive constant. Let

$$\mathbf{w}^{k+1} = \left[\left(\left|s_1^k\right| + \varepsilon\right)^{q-1}, \cdots \left(\left|s_M^k\right| + \varepsilon\right)^{q-1}\right] \quad (13)$$

denote the updated weighting vector based on $\mathbf{s}^k$ of the k-th iteration, then, the iteratively reweighted algorithm update the parameters at the $(k+1)$-th iteration as

$$\underset{\mathbf{x},\mathbf{s}}{\text{minimize}} \left\|\mathbf{w}^{k+1} \odot \mathbf{s}\right\|_1$$
$$\text{subject to} \quad \mathbf{Ax} \le \mathbf{b} + \mathbf{s} \otimes \mathbf{1}_{\kappa \times 1}, \quad \mathbf{s} \ge 0 \quad (14)$$

where $\odot$ denotes the Hadamard product.

Similar to (12), the problem (14) can be solved by the duality approach. The dual problem of (14) is given by

$$\underset{\mathbf{y},\mathbf{v}}{\max} -\mathbf{b}^T \mathbf{y}$$
$$\text{subject to} \quad \mathbf{A}^T \mathbf{y} = \mathbf{0}$$
$$\mathbf{w}^{k+1} - \mathbf{Jy} - \mathbf{v} = \mathbf{0} \quad (15)$$
$$\mathbf{y} \ge \mathbf{0}, \quad \mathbf{v} \ge \mathbf{0}.$$

The iteratively reweighted algorithm is summarized as follows.

Algorithm 2: Iteratively reweighted algorithm

Input: The set of measurements Ω, inlier threshold $\delta > 0$, $\varepsilon > 0$, $q \in (0,1)$, an initialization $\mathbf{w}^1 = \mathbf{1}_{M \times 1}$.
Begin:
    Construct $\mathbf{A}$ and $\mathbf{b}$ from the measurements Ω.
    For $k = 1, 2, \cdots, K$
        Solve (15) with $\mathbf{w}^k$ to obtain $\mathbf{s}^k$.
        Update $\mathbf{w}^{k+1}$ via (13) based on $\mathbf{s}^k$.
    End For
    Set $\mathbf{s} = \mathbf{s}^K$ and remove the residuals for which $s_i > 0$.
End
Output: A subset $I \subseteq \Omega$ of the measurements for which $\exists \mathbf{x}$ such that $E_i(\mathbf{x}) \le \delta$, $\forall i \in I$.

This algorithm iteratively updates the weighting vector and solves the LP problem (15) K times. In practical applications, a small value of K can yield sufficiently good performance, as shown in the experiments.

IV. EXPERIMENTS

This section evaluates the proposed algorithms via experiments in comparison with the following methods:

(a) RANSAC [26]: $\rho = 0.99$ is used for stopping criterion.
(b) Exact penalty (EP) method [24][1]: initialized by the solution of the least squares method. It deterministically

---

[1] Code at: https://www.researchgate.net/publication/320707327_demo_pami.
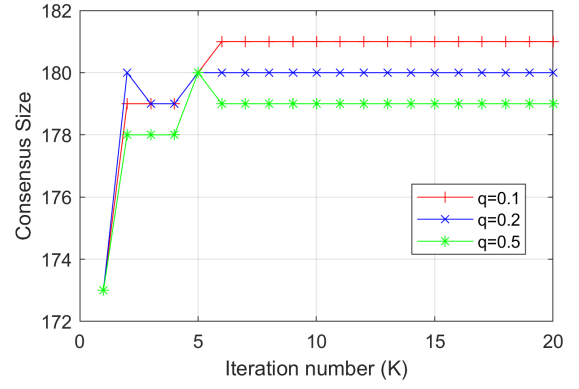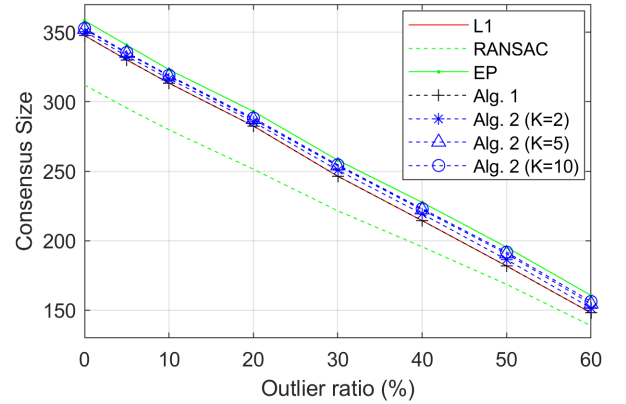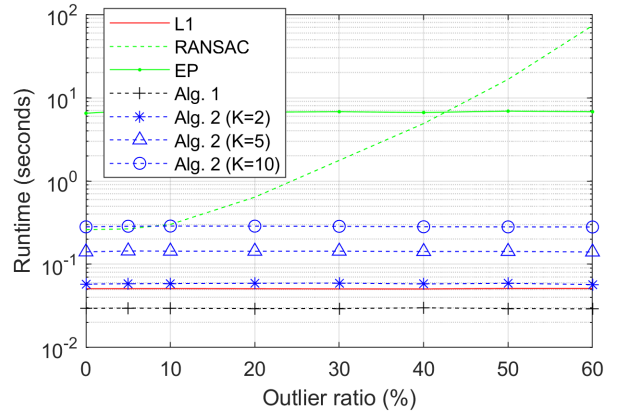

Fig. 1. Results of Algorithm 2 versus iteration number for different values of q (the outlier ratio is 50%).

(a) Number of remaining inliers (consensus size)

(b) Runtime

Fig. 2. Results on robust linear regression versus outlier ratio.

solves a reformulation of the consensus maximization problem with linear complementarity constraints by the Frank-Wolfe algorithm.

(c) ℓ_1 method [6][2]: it solves the dual problem of (10) using an LP solver, which is one of the most efficient algorithms that suitable for large-scale 3D-reconstruction.

---

[2] Code is available at: http://www.maths.lth.se/matematiklth/personal/calle/.

TABLE I
RESULTS ON THE HOUSE DATASET. IT CONTAINS 23 CAMERAS AND 35470 3D-POINTS ARE VISIBLE IN AT LEAST 2 IMAGES. THE RMSE WITHOUT OUTLIER REMOVAL IS 4.76 PIXELS.

	ℓ_∞ [11]	ℓ_1 [6]	Alg. 1	Alg. 2
Removed outliers	2396	862	722	498
Remaining inliers	99936	101470	101610	101834
RMSE (pixels)	0.9119	0.5982	0.6007	0.6063
Runtime (seconds)	50878	975	322	477

TABLE II
RESULTS ON THE CATHEDRAL DATASET. IT CONTAINS 17 CAMERAS AND 16961 3D-POINTS ARE VISIBLE IN AT LEAST 2 IMAGES. THE RMSE WITHOUT OUTLIER REMOVAL IS 3.24 PIXELS.

	ℓ_∞ [11]	ℓ_1 [6]	Alg. 1	Alg. 2
Removed outliers	1594	652	541	428
Remaining inliers	44451	45393	45504	45617
RMSE (pixels)	1.1343	0.8036	0.8075	0.8178
Runtime (seconds)	2354	259	76	134

TABLE III
RESULTS ON THE COLLEGE DATASET. IT CONTAINS 57 CAMERAS AND 8990 3D-POINTS WHICH ARE VISIBLE IN AT LEAST 2 IMAGES. THE RMSE WITHOUT REMOVAL OF ANY OUTLIER IS 4.16 PIXELS.

	ℓ_∞ [11]	ℓ_1 [6]	Alg. 1	Alg. 2
Removed outliers	966	459	319	179
Remaining inliers	26756	27263	27403	27543
RMSE (pixels)	1.0754	0.5134	0.5197	0.5562
Runtime (seconds)	4880	162	51	79

(d) ℓ_∞ method [11]: solved using Gugat's algorithm [22]. This method also use slack variables and minimizes the maximum slack value. It repeatedly removes the data with the largest slack value, until the largest slack value is not greater than zero.

We use $q = 0.1$ and $\varepsilon = 10^{-3}$ for Algorithm 2. SeDuMi [23] is used to solve the involved LP problems in all these algorithms. The experiments were implemented in MATLAB and run on a laptop with 2 GHz Intel I7 CPU and 16 GB RAM.

A. Robust Linear Regression on Synthetic Data

Before proceeding to the main interest of 3D-reconstruction, we repeat a synthetic experiment in [24] on a small linear regression problem with synthetic data. We generated $M = 500$ points $\mathbf{y} = \mathbf{A}\mathbf{x}$ with $\mathbf{A} \in \mathbb{R}^{500 \times 8}$ and $\mathbf{x} \in \mathbb{R}^8$. The elements of $\mathbf{A}$ follow a uniform distribution in $[-1, 1]$. $\mathbf{y}$ is perturbed by white Gaussian noise with standard deviation of 0.1. To simulate outliers, a part of the elements in $\mathbf{y}$ are corrupted by much higher Gaussian noise with standard deviation of 1.

Fig. 1 shows the performance of the iteratively reweighted algorithm (Algorithm 2) versus the iteration number K for different values of q, e.g., $q \in \{0.1, 0.2, 0.5\}$. It can be seen that, Algorithm 2 converges within a few iterations, e.g., $K < 10$. Fig. 2 shows the results of the algorithms for different outlier ratio. Each result is an average over 100 independent

run. Compared with the ℓ_1 method, Algorithm 1 removes a same number of outliers in all cases, while Algorithm 2 removes fewer outliers. Algorithm 2 can achieve sufficiently good performance within a few iterations, e.g. $K = \{2, 5\}$. The EP method yields the largest consensus size. In terms of runtime, Algorithm 1 is the fastest while Algorithm 2 is much faster than EP.

B. Global SfM on Real World Datasets

We consider the full 3D-reconstruction experiments in [6] and use the known-rotation formulation for outlier removal, which is a procedure of RANSAC followed by outlier removal and bundle adjustment [18]. The initial camera rotation and image data are obtained using RANSAC for pairwise images. The global outlier removal is achieved over the structure and translation of the cameras. The inlier threshold is selected corresponding to a reprojection error tolerance of 5 pixels.

Three open datasets[3] are used, including a house (consists of 23 cameras and 29220 3D points projected into 35470 image points), a cathedral (consists of 17 cameras and 16961 3D points projected into 46045 image points), and a college (consists of 57 cameras and 8990 3D points projected into 27722 image points). In these datasets, SIFT descriptors [17] are used to generate point correspondences and, then, RANSAC is used to discard outliers and determine orientations between pairs of cameras.

Since RANSAC and EP are not suitable for large-scale problems, only the ℓ_1 [6] and ℓ_∞ [11] methods are compared here. We use $q = 0.1$ and $K = 2$ for Algorithm 2. Fig. 3 illustrates the resulting reconstruction on the three datasets. Table 1 compares the reconstruction results, including removed outliers, remaining inliers, root-mean-squared-error (RMSE) of reprojection in pixels, and runtime in seconds. Without removal of any outliers, the reprojection RMSEs in reconstructing the house, cathedral and college are respectively 4.76, 3.24, and 4.16 pixels. It can be seen that each algorithm can significantly reduce the RMSE, which implies the effective removal of some outliers. A significant drop in the reprojection RMSE implies the removed data are true outliers.

Algorithm 1 is more than three times faster than the ℓ_1 method, which is due to the reduced dimension. For example, in the cathedral experiment, the number of variables in the ℓ_1 method is 327201 ($\mathbf{x} \in \mathbb{R}^{50931}$ contains the parameters of the 3D points and camera translations, and $\tilde{\mathbf{s}} \in \mathbb{R}^{276270}$ contains the slack variables). Whereas, in our formulation the number of variables is 96976, with $\mathbf{x} \in \mathbb{R}^{50931}$ and $\mathbf{s} \in \mathbb{R}^{46045}$. In the experiments, $\kappa = 6$ as a depth constraint is also considered.

Compared with the ℓ_1 method, Algorithm 1 removes fewer outliers while yields a comparable RMSE. The removal of fewer outliers is in contrast to the results in the linear regression experiment with synthetic data. Algorithm 2 removes much fewer outliers compared with the other algorithms, while being much faster than the ℓ_1 and ℓ_∞ methods.

---

[3] Available online at http://www.maths.lth.se/matematiklth/personal/calle/.

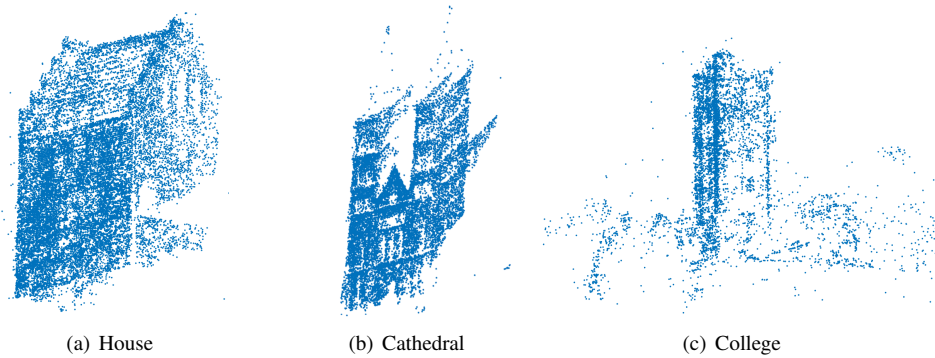|		
(a) House	(b) Cathedral	(c) College

Fig. 3. The resulting reconstruction on the three datasets.

V. CONCLUSION

In Global Structure-from-Motion, feature point matching often generates some mismatches, which gives rise to outliers. This work developed two efficient methods to detect and remove such outliers. Compared with existing methods, the new methods use a dimension reduced formulation, which significantly reduces the computational complexity. Realistic multiview reconstruction experiments demonstrated that, the new algorithms are much faster than state-of-the-art algorithms (e.g., more than three times faster than the ℓ_1 method) while give an improved solution. Due to their efficiency and effectiveness, the new methods could be useful in practical large-scale Structure-from-Motion applications.

REFERENCES

[1] P. Meer, "Robust techniques for computer vision," in G. Medioni and S. B. Kang eds.: *Emerging topics in computer vision*. Prentice Hall, pp. 107–190, 2004.

[2] F. Wen, R. Ying, Z. Gong, and P. Liu, "Efficient algorithms for maximum consensus robust fitting," *IEEE Trans. Robotics*, 2019.

[3] D. Zou and P. Tan, "Coslam: Collaborative visual slam in dynamic environments," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 354–366, 2012.

[4] H. Le, T. J. Chin, and D. Suter, "An exact penalty method for locally convergent maximum consensus," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1888–1896.

[5] T. J. Chin, Y. H. Kee, A. Eriksson, and F. Neumann, "Guaranteed outlier removal with mixed integer linear programs," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5858–5866.

[6] C. Olsson, A. P. Eriksson, and R. Hartley, "Outlier removal using duality," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2010 pp. 1450–1457.

[7] F. Kahl and R. Hartley, "Multiple view geometry under the ℓ_∞-norm," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 9, pp. 1603–1617, 2008.

[8] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp.73–101, 1964.

[9] S. Agarwal, N. Snavely, and S. Seitz, "Fast algorithms for ℓ_∞ problems in multiview geometry," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Anchorage, USA, 2008.

[10] K. Astrom, O. Enqvist, C. Olsson, and F. Kahl, "An ℓ_∞ approach to structure and motion problems in 1d-vision," in *Int. Conf. Computer Vision*, Rio de Janeiro, Brazil, 2007.

[11] K. Sim and R. Hartley, "Removing outliers using the l-infinity norm," in *Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 485–494, New York City, USA, 2006.

[12] F. Wen, R. Ying, P. Liu, and R. C. Qiu, "Robust PCA Using Generalized Nonconvex Regularization," *IEEE Trans. Circuits Systems Video Technology*, DOI:10.1109/TCSVT.2019.2908833, 2019.

[13] E. J. Candes, M. B. Wakin, S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008

[14] Y. Seo, H. Lee, and S. W. Lee, "Outlier removal by convex optimization for l-infinity approaches," in *PSIVT09: Pacific Rim Symposium on Advances in Image and Video Technology*, 2009.

[15] A. Dalalyan and R. Keriven, "L1-penalized robust estimation for a class of inverse problems arising in multiview geometry," in *Annual Conference on Neural Information Processing Systems*, Vancouver, Canada, 2009.

[16] Q. Ke and T. Kanade, "Quasiconvex optimization for robust geometric reconstruction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1834–1847, 2007.

[17] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal Computer Vision*, 2004.

[18] D. Martinec and T. Pajdla, "Robust rotation and translation estimation in multiview reconstruction," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Minneapolis, USA, 2007.

[19] C. Olsson, O. Enqvist, and F. Kahl, "A polynomial-time bound for matching and registration with ouliers," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Anchorage, USA, 2008.

[20] F. Wen, R. Ying, P. Liu, and T.K. Trong, "Nonconvex regularized robust PCA using proximal block coordinate descent algorithm," *IEEE Trans. Signal Processing*, DOI:10.1109/TSP.2019.2940121, 2019.

[21] F. Wen, P. Liu. Y. Liu, R. C. Qiu, and W. Yu, "Robust sparse recovery in impulsive noise via Lp-L1 optimization," *IEEE Trans. Signal Processing*, vol. 65, no. 1, pp. 105–118, Jan. 2017.

[22] M. Gugat, "A fast algorithm for a class of generalized fractional programs," *Man. Sci.*, vol. 42, no. 10, pp. 1493–1499, 1996.

[23] J. F. Sturm. *Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones*. Optimization Methods and Software, vol. 11–12, pp. 625–653, 1999.

[24] H. Le, T. J. Chin, A. Eriksson, and D. Suter, "Deterministic approximate methods for maximum consensus robust fitting," *arXiv preprint*, arXiv:1710.10003, 2017.

[25] T. J. Chin, P. Purkait, A. Eriksson, and D. Suter, "Efficient globally optimal consensus maximisation with tree search," in *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 2413–2421.

[26] 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.

[27] O. Chum, J. Matas, and J. Kittler, "Locally optimized ransac," in *DAGM*. Springer, 2003.

[28] Z. Cui and P. Tan, "Global structure-from-motion by similarity averaging," in *Proc. IEEE Int. Conf. Computer Vision*, 2015.

[29] Z. Gong, R. Ying, F. Wen, and P. Liu, "Tightly coupled integration of GNSS and vision SLAM using 10-DoF optimization on manifold," *IEEE Sensors Journal*, DOI:10.1109/JSEN.2019.2935387, 2019.

[30] P. Purkait, C. Zach, and A. Eriksson, "Maximum consensus parameter estimation by reweighted L1 methods," in *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2018, pp. 312–32.