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 $\ell_1$ minimization and is solved by a single linear programming (LP). The second method approximately solves the ideal $\ell_0$ 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 $\ell_1$ 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]. Given a set of $M$ measurements $\{a_i, y_i\}_{i=1,\cdots,M}$, 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 $x \in \mathbb{R}^N$, that is consistent with as many of the input data as possible, i.e., has the largest consensus set $I$ as

$$\maximize_{I \subseteq \Omega} |I|$$
subject to $|a_i^T x - y_i| \leq \delta, \quad \forall i \in I$  \hspace{1cm} (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 $\Omega \setminus I^*$ denotes the index set of the true outliers.

On the one hand, due to the intractability of the robust geometric fitting problem, 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 [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. The most efficient algorithm suitable for this problem is the $\ell_1$ method [6], which solves a convex relaxation of (1). Although there is no theoretical guarantee of success, the $\ell_1$ 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 $\ell_1$ 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 $\ell_1$ minimization that is solved by linear
programming (LP). The second one approximately solves the ideal $\ell_0$ 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 $s \in \mathbb{R}^M, s \geq 0$, problem (1) can be recast into the following formulation

$$\begin{align*}
\text{maximize } & \quad M - \|s\|_0 \\
\text{subject to } & \quad |\alpha_i^T x - y_i| \leq \delta + s_i, \quad s_i \geq 0
\end{align*}$$

where $\|s\|_0$ denotes the $\ell_0$ norm which counts the number of nonzero elements in $s$. For a solution $s^*$ of (2), it holds $\text{supp}(s^*) = \Omega_i T^*$, which is the index set of the outliers, and $M - \|s\|_0 = |T^*|$. $\text{supp}(\cdot)$ denotes the support set (the index set of nonzero) of a vector. Equivalently, problem (2) can be expressed as the constrained $\ell_0$ minimization

$$\begin{align*}
\text{minimize } & \quad \|s\|_0 \\
\text{subject to } & \quad |\alpha_i^T x - y_i| \leq \delta + s_i, \quad s_i \geq 0
\end{align*}$$

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

$$\begin{align*}
\alpha_i^T x - y_i \leq \delta + s_i \quad \text{and} \quad -\alpha_i^T x + y_i \leq \delta + s_i.
\end{align*}$$

Let $A = [A_1^T, A_2^T, \cdots, A_M^T]^T$ and $b = [b_1^T, b_2^T, \cdots, b_M^T]^T$ with $A_i = [x_i, -x_i]^T$ and $b_i = [y_i + \delta, -y_i + \delta]^T$. Then, the formulation (3) can be rewritten as

$$\begin{align*}
\text{minimize } & \quad \|s\|_0 \\
\text{subject to } & \quad Ax \leq b + s \otimes 1_{\kappa \times 1}, \quad s \geq 0
\end{align*}$$

where $\kappa = 2$, $1_{\kappa \times 1}$ is a $\kappa$-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 $|\alpha^T x - y|$, 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 $(z_1^t, z_2^t)$ be a measurement in one of the images and $z$ be its corresponding 3D-point. Given the camera matrix $P = [R, t] \in \mathbb{R}^{3 \times 4}$, the squared projection error is

$$E_i(z, P) = \left\| z_1^t - \frac{r_1 z + t_1}{r_3 z + t_3}, z_2^t - \frac{r_2 z + t_2}{r_3 z + t_3} \right\|^2_2$$

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

$$E_i(x) = \frac{\|u_i^T x + \tilde{u}_i, v_i^T x + \tilde{v}_i\|_2^2}{(w_i^T x + \tilde{w}_i)^2}.$$  

In many geometry problems in computer vision, if either $z$ or $R$ is known, the residual formulation can be rewritten in the form of (7). For example, for the triangulation problem, $R$ and $t$ are kept fixed, and $x$ 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, $x$ in (7) contains the parameters of the 3D-points and the camera translations, which is also called the known rotation problem.

To handle such geometric residuals, in addition to the Euclidean distance, the coordinate-wise max distance ($\ell_\infty$ norm) and the absolute distance ($\ell_1$ norm) are also popular due to their convenience. Specifically, a general formulation of the error residuals is of the form [7], [8]

$$E_i(x) = \frac{\|u_i^T x + \tilde{u}_i, v_i^T x + \tilde{v}_i\|_p}{w_i^T x + \tilde{w}_i}$$

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

$$\|u_i^T x + \tilde{u}_i, v_i^T x + \tilde{v}_i\|_p \leq \delta(w_i^T x + \tilde{w}_i)$$

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

For the $\ell_1$ norm, i.e., $p = 1$, the constraint becomes

$$|u_i^T x + \tilde{u}_i| + |v_i^T x + \tilde{v}_i| \leq \delta(w_i^T x + \tilde{w}_i)$$

which can be expressed as the following linear inequalities

$$\begin{align*}
(u_i^T + v_i^T) x + \tilde{u}_i + \tilde{v}_i & \leq \delta(w_i^T x + \tilde{w}_i) \\
(u_i^T - v_i^T) x + \tilde{u}_i - \tilde{v}_i & \leq \delta(w_i^T x + \tilde{w}_i) \\
(v_i^T - u_i^T) x - \tilde{u}_i + \tilde{v}_i & \leq \delta(w_i^T x + \tilde{w}_i) \\
-(u_i^T + v_i^T) x - \tilde{u}_i - \tilde{v}_i & \leq \delta(w_i^T x + \tilde{w}_i).
\end{align*}$$

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

$$d_{\text{min}} \leq w_i^T x + \tilde{w}_i \leq d_{\text{max}}$$

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

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

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

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

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

which is equivalent to

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

Using (4), equivalent linear inequalities of the constraint (8) can be derived as follows

$$\begin{align*}
(u_i^T + v_i^T) x + \tilde{u}_i + \tilde{v}_i & \leq \delta(w_i^T x + \tilde{w}_i) \\
v_i^T x + \tilde{v}_i & \leq \delta(w_i^T x + \tilde{w}_i).
\end{align*}$$

The effectiveness of the $\ell_\infty$ 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
\[
\min_{x, \tilde{s}} \sum_{i} \tilde{s}_i \quad \text{subject to} \quad |a_i^T x - y_i| \leq \delta + z_i L, \quad z_i \in \{0, 1\}
\]
where \(L\) is a large positive constant. For a solution \(z^*\) of [9], it is easy to see that supp\((z^*) = \text{supp}(s^*)\), 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., \(\ell_1\) approximation [6], [14], [15]. The most efficient \(\ell_1\) method [6] considers a formulation as
\[
\min_{x, \tilde{s}} 1^T \tilde{s} \quad \text{subject to} \quad A x \leq b + \tilde{s}, \quad \tilde{s} \geq 0.
\]
This \(\ell_1\) minimization formulation can be viewed as a convex relaxation of (5), where the nonconvex \(\ell_0\) norm is replaced by its convex envelop, i.e., the \(\ell_1\) norm. Meanwhile, the structure in \(s \otimes 1_{\kappa \times 1}\) in the constraint is ignored and, hence, the number of unknown parameters in \(\tilde{s}\) is \(\kappa M\) while that in \(s\) 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 result.

The \(\ell_1\) 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 \(\ell_0\) or \(\ell_q\) (\(0 < q < 1\)) norm can usually yield a sparser solution than the \(\ell_1\) norm [12]. Empirical results have shown that the \(\ell_1\) minimization (10) is likely to remove inliers in some conditions [6], [24]. Since the \(\ell_0\) and \(\ell_q\) norm penalties tend to yield a sparser solution, they can be expected to reduce the possibility of removing true inliers.

III. Proposed Algorithms

Generally, it is difficult to directly solve the nonconvex \(\ell_0\) 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 \(\ell_0\) minimization problem (5).

A. \(\ell_1\) Algorithm with Reduced Dimension

We consider a convex relaxation of (5) via replacing the \(\ell_0\)-norm by its convex envelop, the \(\ell_1\)-norm, as
\[
\min_{x, \tilde{s}} \|s\|_1 \quad \text{subject to} \quad A x \leq b + s \otimes 1_{\kappa \times 1}, \quad s \geq 0.
\]
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 \(\ell_1\) algorithm [6] is \(\kappa M\). The dimension reduction would result in a speedup. The dual problem of (11) is given by
\[
\max_{y, v} -b^T y \quad \text{subject to} \quad A^T y = 0, \quad 1_{M \times 1} - J y - v = 0, \quad y \geq 0, \quad v \geq 0
\]
where \(y \in \mathbb{R}^{\kappa M}\) and \(v \in \mathbb{R}^M\) are the dual variables, and \(J = I_M \otimes 1_{\kappa \times 1} \in \mathbb{R}^{M \times (\kappa M)}\). \(I_M\) 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: \(\ell_1\) algorithm with reduced dimension

Input: The set of measurements \(\Omega\), inlier threshold \(\delta > 0\).

Begin:
- Construct \(A\) and \(b\) from the measurements \(\Omega\).
- Solve the LP problem (12) to obtain \(s\).
- Remove the residuals for which \(s_i > 0\).

End

Output: A subset \(I \subseteq \Omega\) of the measurements for which \(\exists x\) such that \(E_i(x) \leq \delta, \forall i \in I\).

B. Iteratively Reweighted Algorithm

As shown in [6], [24], the \(\ell_1\) method probably removes true inliers in practical applications. This is explained in the last section, as the \(\ell_1\) minimization may yield a solution not sparse enough. Ideally, we should minimize the number of nonzero elements in \(s\), i.e., the \(\ell_0\) 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 \(\ell_0\) norm by the \(\ell_q\) norm with a small value of \(q\) (e.g., \(q = 0.1\) in the experiments). The \(\ell_q\) norm usually yields a sparser solution than \(\ell_1\) norm [20], [21]. Then, at the \((k + 1)\)-th iteration, the \(\ell_q\) norm is approximated via first-order expansion (linearization) at \(s^k\) obtained at the \(k\)-th iteration as
\[
\|s\|_q \approx \sum_{i=1}^{M} (|s_i^k| + \varepsilon)^{q-1} |s_i|
\]
where \(\varepsilon\) is a small positive constant. Let
\[
w^{k+1} = \left(\left(|s_1^k| + \varepsilon\right)^{q-1}, \ldots, \left(|s_M^k| + \varepsilon\right)^{q-1}\right)
\]
denote the updated weighting vector based on \( s^k \) of the \( k \)-th iteration, then, the iteratively reweighted algorithm update the parameters at the \((k + 1)\)-th iteration as

\[
\begin{align*}
\text{minimize}_{x,s} & \quad \| w^{k+1} \odot s \|_1 \\
\text{subject to} & \quad Ax \leq b + s \odot 1_{\times 1}, \quad s \geq 0
\end{align*}
\]  

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

\[
\begin{align*}
\max_{y,v} & \quad -b^T y \\
\text{subject to} & \quad A^T y = 0 \\
& \quad w^{k+1} - J y - v = 0 \\
& \quad y \geq 0, \quad v \geq 0.
\end{align*}
\]

(15)

The iteratively reweighted algorithm is summarized as follows.

**Algorithm 2:** Iteratively reweighted algorithm

**Input:** The set of measurements \( \Omega \), inlier threshold \( \delta > 0 \), \( \varepsilon > 0 \), \( q \in (0,1) \), an initialization \( w^1 = 1_{M \times 1} \).

**Begin:**

Construct \( A \) and \( b \) from the measurements \( \Omega \).

For \( k = 1, 2, \cdots, K \)

- Solve (15) with \( w^k \) to obtain \( s^k \).
- Update \( w^{k+1} \) via (13) based on \( s^k \).

**End For**

Set \( s = s^K \) and remove the residuals for which \( s_i > 0 \).

**End**

**Output:** A subset \( I \subseteq \Omega \) of the measurements for which \( \exists x \) such that \( E_i(x) \leq \delta, \forall i \in I \).

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] initialized by the solution of the least squares method. It deterministically solves a reformulation of the consensus maximization problem with linear complementarity constraints by the Frank-Wolfe algorithm.

(c) \( \ell_1 \) method [6] 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.

(d) \( \ell_\infty \) 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 \( y = Ax \) with \( A \in \mathbb{R}^{500 \times 8} \) and \( x \in \mathbb{R}^8 \). The elements of \( A \) follow a uniform distribution in \([-1,1] \). \( y \) is perturbed
TABLE I
Reconstruction results on the House dataset. It contains 23 cameras and 35470 3D-points which are visible in at least 2 images. The RMSE without removal of any outlier is 4.76 pixels.

|        | \(\ell_\infty\) (Gugat’s) [11] | \(\ell_1\) [6] | Alg. 1 | Alg. 2 \((K = 2)\) |
|--------|---------------------------------|----------------|-------|------------------|
| House  | Removed outliers                | 2396           | 862   | 722              |
|        | Remaining inliers                | 99936          | 101470| 101610           |
|        | RMSE (pixels)                   | 0.9119         | 0.5982| 0.6007           |
|        | Runtime (seconds)               | 50878          | 975   | 322              |

TABLE II
Reconstruction results on the Cathedral dataset. It contains 17 cameras and 16961 3D-points which are visible in at least 2 images. The RMSE without removal of any outlier is 3.24 pixels.

|        | \(\ell_\infty\) (Gugat’s) [11] | \(\ell_1\) [6] | Alg. 1 | Alg. 2 \((K = 2)\) |
|--------|---------------------------------|----------------|-------|------------------|
| Cathedral | Removed outliers                | 1594           | 652   | 541              |
|        | Remaining inliers                | 44451          | 45393 | 45504            |
|        | RMSE (pixels)                   | 1.1343         | 0.8036| 0.8075           |
|        | Runtime (seconds)               | 2354           | 259   | 76               |

TABLE III
Reconstruction 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.

|        | \(\ell_\infty\) (Gugat’s) [11] | \(\ell_1\) [6] | Alg. 1 | Alg. 2 \((K = 2)\) |
|--------|---------------------------------|----------------|-------|------------------|
| College | Removed outliers                | 966            | 459   | 319              |
|        | Remaining inliers                | 26756          | 27263 | 27403            |
|        | RMSE (pixels)                   | 1.0754         | 0.5134| 0.5197           |
|        | Runtime (seconds)               | 4880           | 162   | 51               |

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

by white Gaussian noise with standard deviation of 0.1. To simulate outliers, a part of the elements in \(y\) 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 \(\ell_1\) 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 are used, including a house (consists of

\footnote{Available online at \url{http://www.maths.lth.se/matematiklth/personal/calle/}.}
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 $\ell_1$ [6] and $\ell_\infty$ [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 $\ell_1$ method, which is due to the reduced dimension. For example, in the cathedral experiment, the number of variables in the $\ell_1$ method is 327201 ($x \in \mathbb{R}^{50931}$ contains the parameters of the 3D points and camera translations, and $s \in \mathbb{R}^{276270}$ contains the slack variables). Whereas, in our formulation the number of variables is 96976, with $x \in \mathbb{R}^{50931}$ and $s \in \mathbb{R}^{46045}$. In the experiments, $\kappa = 6$ as a depth constraint is also considered.

Compared with the $\ell_1$ 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 $\ell_1$ and $\ell_\infty$ methods.

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 $\ell_1$ method) while give 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] R. Mur-Artal, J. M. M. Montiel, and J. D Tardós, “Osb-slam: a ver-satile and accurate monocular slam system,” IEEE Trans. Robotics, vol. 31, no. 5, pp. 1147–1163, 2015.
[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 $\ell_\infty$-norm,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, no. 9, pp. 1603–1617, 2008.
[8] P. Meer, “Robust techniques for computer vision,” in G. Medioni and L. Zelnik-Manor (eds.), Emerging topics in computer vision, 2004.
[9] S. Agarwal, N. Snavely, and S. Seitz, “Fast algorithms for $\ell_\infty$ problems in multiview geometry,” in IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Anchorage, USA, 2008.
[10] K. Astrom, O. Engqvist, C. Olsson, and F. Kahl, “An $\ell_\infty$ approach to structure and motion problems in 1d-motion,” in Int. Conf. Computer Vision, Rio de Janeiro, Brazil, 2007.
[11] K. Sim and R. Hartley, “Removing outliers using the $\ell_1$-infinity norm,” in IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 485–494, New York City, USA, 2006.
[12] F. Wen, P. Liu, and R. Qiu, “Nonconvex regularization based sparse and low-rank recovery in signal processing, statistics, and machine learning,” arXiv preprint arXiv:1808.05403, 2018.
[13] E. J. Candes, M. B. Wakin, S. P. Boyd, “Enhancing sparsity by reweighted 11 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 $\ell_1$-infinity approaches,” in PSIVT09: Pacific Rim Symposium on Advances in Image and Video Technology, 2009.
[15] D. Lowe, “Distinctive image features from scale-invariant keypoints,” Int. Journal Computer Vision, 2004.
[16] D. Martinez and T. Pajdla, “Robust rotation and translation estimation in multiview reconstruction,” in IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Minneapolis, USA, 2007.
[17] C. Olsson, O. Engqvist, and F. Kahl, “A polynomial-time bound for matching and registration with outliers,” in IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Anchorage, USA, 2008.

[18] F. Wen, L. Pei, Y. Yang, W. Yu, and P. Liu, “Efficient and robust recovery of sparse signal and image using generalized nonconvex regularization,” IEEE Trans. Computational Imaging, vol. 3, no. 4, pp. 566–579, 2017.
[19] 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 Process., vol. 65, no. 1, pp. 105–118, Jan. 2017.
[20] M. Gutg, “A fast algorithm for a class of generalized fractional programs,” Mar. Sci., vol. 42, no. 10, pp. 1493–1499, 1996.
[21] 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.
[22] H. Le, T. J. Chin, A. Eriksson, and D. Suter, “Deterministic approximate methods for maximum consensus robust fitting,” arXiv preprint arXiv:1710.10003, 2017.
[23] T. J. Chin, P. Parkait, A. Eriksson, and D. Suter, “Efficient globally optimal consensus maximisation with true search,” in IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015, pp. 2413–2421.
[24] M. A. Fischer 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.
[25] D. Cham, J. Matas, and J. Kittler, “Locally optimized ransac,” in DAGM. Springer, 2003.
[26] Z. Cui, and P. Tan, “Global structure-from-motion by similarity averaging,” in Proc. of the IEEE Int. Conf. Computer Vision, 2015.