



Fast approximate L_∞ minimization: Speeding up robust regression



Fumin Shen^{a,b,c}, Chunhua Shen^{c,*}, Rhys Hill^c, Anton van den Hengel^c, Zhenmin Tang^b

^a School of Computer Science and Engineering, University of Electronic Science and Technology of China, China

^b School of Computer Science and Technology, Nanjing University of Science and Technology, China

^c School of Computer Science, The University of Adelaide, Australia

ARTICLE INFO

Article history:

Received 29 July 2012

Received in revised form 24 February 2014

Accepted 24 February 2014

Available online 5 March 2014

Keywords:

Least-squares regression

Outlier removal

Robust regression

Face recognition

ABSTRACT

Minimization of the L_∞ norm, which can be viewed as approximately solving the non-convex least median estimation problem, is a powerful method for outlier removal and hence robust regression. However, current techniques for solving the problem at the heart of L_∞ norm minimization are slow, and therefore cannot be scaled to large problems. A new method for the minimization of the L_∞ norm is presented here, which provides a speedup of multiple orders of magnitude for data with high dimension. This method, termed *Fast L_∞ Minimization*, allows robust regression to be applied to a class of problems which was previously inaccessible. It is shown how the L_∞ norm minimization problem can be broken up into smaller sub-problems, which can then be solved extremely efficiently. Experimental results demonstrate the radical reduction in computation time, along with robustness against large numbers of outliers in a few model-fitting problems.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Linear least-squares (LS) estimation, or linear L_2 norm minimization (denoted as L_2 for brevity throughout this paper), is widely used in computer vision and image analysis due to its simplicity and efficiency. Recently the L_2 norm technique has been applied to recognition problems such as face recognition (Naseem et al., 2010; Shi et al., 2011). All of these methods are linear regression-based and the regression residual is utilized to make the final classification. However, a small number of outliers can drastically bias L_2 , leading to low quality estimates. Clearly, robust regression techniques are critical when outliers are present.

The literature contains a range of different approaches to robust regression. One commonly used method is the M -estimator framework (Huber, 1973, 1981), where the Huber function is minimized, rather than the conventional L_2 norm. Related methods include L -estimators (Koenker and Portnoy, 1987) and R -estimators (Jureckova, 1971). One drawback of these methods is that they are still vulnerable to bad leverage outliers (Hubert et al., 2008). By bad leverage points, we mean those observations that are outlying in x -space and do not follow the linear pattern of the majority. Least median of squares (LMS) (Rousseeuw, 1984), least trimmed squares (LTS) (Rousseeuw and Leroy, 1987) and the technique using data partition and M -estimation (Park et al., 2012) have high-breakdown points. Although each of these regression methods is, in general, more robust than L_2 , they have rarely been applied to object recognition problems in computer vision due to their computational expense (Naseem et al., 2012; Fidler et al., 2006).

* Corresponding author. Tel.: +61 883136745.

E-mail addresses: chhshen@gmail.com, chunhua.shen@adelaide.edu.au (C. Shen).

Another class of robust methods has been developed to remove these abnormal observations from the measurement data. One of the most popular methods is RANSAC (Fischler and Bolles, 1981) which attempts to maximize the size of the *consensus set*. RANSAC relies on iterative random sampling and consensus testing where the size of each sample is determined by the minimum number of data points required to compute a single solution. RANSAC's efficiency is therefore directly tied to the time and number of data points required to compute a solution. For example, RANSAC has been successfully applied to multiview structures-from-motion and homography to estimation problems. However, it is unclear how to apply RANSAC to visual recognition problems, e.g., face recognition, where face images are usually in a high-dimensional space.

Sim and Hartley (2006) proposed an outlier-removing method using the L_∞ norm, which iteratively fits a model to the data and removes the measurement with the largest residual at each iteration. Generally, the iterative method can fail for the L_2 optimization problems, however it is valid for a wide class of L_∞ problems. Sim and Hartley proved that the set of measurements with largest residual must contain at least one outlier. Hence continuing to iterate eventually removes all the outliers. This method is shown to be effective in outlier detection for multiview geometry problems (Sim and Hartley, 2006; Olsson et al., 2010).

L_∞ norm minimization can be time-consuming, since at each step one needs to solve an optimization problem via Second-Order Cone Programming (SOCP) or Linear Programming (LP) in the application of multi-view geometry (Kahl, 2005; Ke and Kanade, 2005). The software package SeDuMi (Sturm, 1999) provides solvers for both SOCP and LP problems.

In this paper, we propose a fast algorithm to minimize the L_∞ norm for approximating the least median estimation (denoted as L_∞ for brevity throughout the paper). Observing that the L_∞ norm is determined by only a few measurements, the optimization strategy *column generation* (CG) (Lübbecke and Desrosiers, 2005) can be applied to reduce the main problem into a set of much smaller sub-problems. Each sub-problem can be formulated as a Quadratic Programming (QP) problem. Due to its relatively small size, the QP problem can be solved extremely efficiently using customized solvers. In particular, we can generate solvers using the technique introduced by Mattingley and Boyd (2012). This reduction results in a speedup of several orders of magnitude for high dimensional data.

This degree of speedup allows L_∞ to be applied to problems which were previously inaccessible. We show how the L_∞ outlier removal technique can be applied to several classification problems in computer vision. Representations of objects in this type of problems are often derived by using L_2 to solve equations containing samples in the same class (or other collaborative classes) (Naseem et al., 2010; Shi et al., 2011). Representation errors are then taken into classification where the query object is assigned to the class corresponding to the minimal residual. This method is shown to be effective on data without occlusion or outliers. However, in real-world applications, measurement data are almost always contaminated by noises or outliers. Before a robust representation can be obtained via linear estimation, outlier removal is necessary. The proposed method is shown to significantly improve the classification accuracies in our experiments for face recognition and iris recognition on several public datasets.

2. Related work

Hartley and Schaffalitzky (2004) seek a globally optimal solution for multi-view geometry problems via L_∞ norm optimization, based on the fact that many geometry problems in computer vision have a single local, and hence global, minimum under the L_∞ norm. In contrast, the commonly used L_2 cost function typically has multiple local minimum (Hartley and Schaffalitzky, 2004; Sim and Hartley, 2006). This work has been extended by several authors, yielding a large set of geometry problems whose globally optimal solution can be found using the L_∞ norm (Olsson provides a summary (Olsson et al., 2007)).

It was observed that these geometry problems are examples of *quasiconvex* optimization problems, which are typically solved by a sequence of SOCPs using a bisection (binary search) algorithm (Kahl, 2005; Ke and Kanade, 2005). Olsson et al. (2007) show that the functions involved in the L_∞ norm problems are in fact *pseudoconvex* which is a stronger condition than quasiconvex. As a consequence, several fast algorithms have been proposed (Olsson et al., 2007; Li, 2009).

Sim and Hartley (2006) propose a simple method based on the L_∞ norm for outlier removal, where measurements with maximal residuals are thrown away. The authors prove that at least one outlier is removed at each iteration, meaning that all outliers will be rejected in a finite number of iterations. However the method is not efficient, since one needs to solve a sequence of SOCPs. Observing that many fixed-dimensional L_∞ geometry problems are actually instances of LP-type problem, an LP-type framework was proposed for the multi-view triangulation problem with outliers (Li, 2007).

Recently, the Lagrange dual problem of the L_∞ minimization problem posed in Hartley and Schaffalitzky (2004) was derived in Olsson et al. (2010). To further boost the efficiency of the method, the authors Olsson et al. (2010) proposed an L_1 -minimization algorithm for outlier removal. While the aforementioned methods add a single slack variable and repeatedly solve a feasibility problem, the L_1 algorithm adds one slack variable for each residual and then solves a single convex program. Though efficient, this method is only successful on data drawn from particular statistical distributions.

Robust statistical techniques, including the aforementioned robust regression and outlier removal methods, can significantly improve the performance of their classic counterparts. However, they have rarely been applied in the image analysis field, to problems such as visual recognition, due to their computational expense. The M -estimator method is utilized in Naseem et al. (2012) for face recognition and achieved high accuracy even when illumination change and pixel corruption were present. In Fidler et al. (2006), the authors propose a theoretical framework combining reconstructive and

discriminative subspace methods for robust classification and regression. This framework acts on subsets of pixels in images to detect outliers.

The remainder of this paper is organized as follows. In Section 3 we briefly review the L_2 and L_∞ problems. In Section 4, the main outlier removal algorithm is presented. In Section 5, we formulate the L_∞ norm minimization problem into a set of small sub-problems which can be solved with high efficiency. We then apply the outlier removal technique in Section 6 to several visual recognition applications. Finally the conclusion is given in Section 8.

3. The L_2 and L_∞ norm minimization problems

In this section, we briefly present the L_∞ norm minimization problem in the form we use in several recognition problems. Let us first examine the L_2 norm minimization problem,

$$\min_{\beta} \sum_{i=1}^n (\mathbf{a}_i \beta - y_i)^2, \tag{1}$$

for which we have a closed-form solution¹

$$\beta_{ls}^* = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y}, \tag{2}$$

where $\mathbf{A} = [\mathbf{a}_1^\top; \dots; \mathbf{a}_n^\top] \in \mathbb{R}^{n \times d}$ is the measurement data matrix, composed of rows $\mathbf{a}_i \in \mathbb{R}^d$, and usually $n \gg d$. The model's response is represented by the vector $\mathbf{y} \in \mathbb{R}^n$ and $\beta \in \mathbb{R}^d$ stores the parameters to be estimated. Note that in our visual recognition applications, both \mathbf{y} and the columns of \mathbf{A} are images flattened to vectors. According to the linear subspace assumption (Basri and Jacobs, 2003), a probe image can be approximately represented by a linear combination of the training samples of the same class: $\mathbf{y} \approx \mathbf{A}\beta$. Due to its simplicity and efficacy, the linear representation method is widely used in various image analysis applications, e.g., Naseem et al. (2010); Shi et al. (2011); Zhang et al. (2011); Wright et al. (2009); Yang et al. (2012); Pillai et al. (2011).

The L_2 norm minimization aims to minimize the sum of squared residuals (1), where the terms $f_i(\beta) = (\mathbf{a}_i \beta - y_i)^2$, $i \in I = \{1, \dots, n\}$, are the squared residuals. L_2 norm minimization is simple and efficient, however it utilizes the entire dataset and therefore can be easily influenced by outliers.

Instead of minimizing the sum of squared residuals, the L_∞ - L_2 norm minimization method seeks to minimize only the maximal residual, leading to the following formulation:

$$\min_{\beta} \max_i (\mathbf{a}_i \beta - y_i)^2, \quad i \in I. \tag{3}$$

This equation has no closed-form solution, however it can be easily reformulated into a constrained formulation, with an auxiliary variable:

$$\begin{aligned} \min_{\beta} \quad & s \\ \text{s.t.} \quad & (\mathbf{a}_i \beta - y_i)^2 \leq s, \quad \forall i \in I; \end{aligned} \tag{4}$$

which is clearly a SOCP problem. If we take the absolute value of the residual in (3), we obtain

$$\min_{\beta} \max_i |\mathbf{a}_i \beta - y_i|, \quad i \in I \tag{5}$$

leading to an LP problem

$$\begin{aligned} \min_{\beta} \quad & s \\ \text{s.t.} \quad & \mathbf{a}_i \beta - y_i \leq s, \\ & \mathbf{a}_i \beta - y_i \geq -s, \quad \text{for all } i \in I. \end{aligned} \tag{6}$$

A critical advantage of the L_∞ norm cost function is that it has a single global minimum in many multi-view geometry problems (Hartley and Schaffalitzky, 2004; Li, 2007). Unfortunately, like L_2 norm minimization, the L_∞ norm minimization method is also vulnerable to outliers. Moreover, minimizing the L_∞ norm fits to the outliers, rather than the data truly generated by the model (Sim and Hartley, 2006). Therefore, it is necessary to first reject outliers before the estimation.

4. Outlier removal via maximum residual

In Sim and Hartley (2006), outlier removal is conducted in an iterative fashion by first minimizing the L_∞ norm, then removing the measurements with maximum residual and then repeating the process. The measurements with maximum residual are referred to as the *support set* of the minimax problem, i.e.,

$$I_{supp} = \{i \in I \mid f_i(\beta^*) = \delta_{opt}\}, \tag{7}$$

¹ The closed-form solution can only be obtained when \mathbf{A} is over-determined, i.e., $n \geq d$. When $n < d$, one can solve the multicollinearity problem by ridge regression, or another variable selection method, to obtain a unique solution.

Algorithm 1 Outlier removal using the L_∞ norm

-
- 1: **Input:** the measurement data matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$; the response vector $\mathbf{y} \in \mathbb{R}^n$; outlier percent p .
 - 2: **Initialization:** $l \leftarrow 0$; number of measurements to be removed $L \leftarrow \lfloor n \times p \rfloor$; index $I_t \leftarrow \{1, \dots, n\}$.
 - 3: **while** $l < L$ **do**
 - 4: Solve the L_∞ norm minimization problem: $\delta_{opt} = \min_{\beta} \max_{i \in I_t} f_i(\beta)$; get the support set I_{supp} via equation (7).
 - 5: Remove the measurements with indices in I_{supp} , i.e., $I_t \leftarrow I_t \setminus I_{supp}$.
 - 6: **Remedy.** Solve the minimax problem again with the new index I_t and get the optimal residual δ_{opt} and parameter β^* . Move the indices $I_r = \{i \in I_{supp} | f_i(\beta^*) < \delta_{opt}\}$ back to I_t .
 - 7: $l \leftarrow l + |I_{supp}|$.
 - 8: **end while**
 - 9: **Output:** \mathbf{A}_l and \mathbf{y}_l with measurement index $i \in I_t$.
-

where $\delta_{opt} = \min_{\beta} \max_{i \in I} f_i(\beta)$ is the optimum residual of the minimax problem. The outlier removal strategy does not work well for the general L_2 minimization problems (i.e., $\delta_{opt} = \min_{\beta} \sum_i f_i(\beta)$), because the outliers are not guaranteed to be included in the support set. In contrast, this strategy is valid for the L_∞ minimization problems. For problem (3) or (5), it is proved by the following theorems that the measurements with largest residual must contain at least one outlier.

Suppose the index vector l is composed of l_{in} and l_{out} , the inlier and outlier sets respectively, and there exists δ_{in} such that $\min_{\beta} \max_{i \in l_{in}} f_i(\beta) < \delta_{in}$. Then Sim and Hartley (2006) proved the following theorem.

Theorem 1 (Sim and Hartley, 2006). Consider the L_∞ norm minimization problem (3) or (5) with the optimal residual $\delta_{opt} = \min_{\beta} \max_{i \in I} f_i(\beta)$. If there exists an inlier subset $l_{in} \subset I$ for which $\min_{\beta} \max_{i \in l_{in}} f_i(\beta) < \delta_{in} < \delta_{opt}$, then the support set I_{supp} must contain at least one index $i \in l_{out}$, that is, an outlier.

To prove Theorem 1, we need the following theorem.

Theorem 2. If i_0 is not in the support set I_{supp} for the minimax problem (3) or (5), then removing the measurement with respect to i_0 will not decrease the optimal residual δ_{opt} . Formally, if $i_0 \notin I_{supp}$, then

$$\min_{\beta} \max_{i \in I \setminus i_0} f_i(\beta) = \min_{\beta} \max_{i \in I} f_i(\beta) = \delta_{opt}. \quad (8)$$

Proof. It is not difficult to verify that both the residual error functions $(\mathbf{a}_i \beta - y_i)^2$ and $|\mathbf{a}_i \beta - y_i|$ are convex, and therefore also quasiconvex (Boyd and Vandenberghe, 2004). Furthermore these two error functions are also strictly convex then also strictly quasiconvex. Then due to Corollary 1 in Sim and Hartley (2006) (omitted here), the theorem holds. \square

Based on Theorems 1 and 2 can be easily proved in a similar way as the proof of Theorem 2 in Sim and Hartley (2006). Theorem 2 indicates that the optimal solution of the minimax problem is only determined by the support set. This inspires us to develop a fast algorithm for the minimax problems in the Section 5.

From Theorem 1, we know that by solving the minimax problem (3) or (5) one time, at least one outlier will be removed through excluding the elements corresponding to the support set. Based on that, we summarize our iterative outlier removal method in Algorithm 1. At each iteration, we first obtain the optimal parameters β^* by solving the minimax problem and then remove the measurements (pixels in images) corresponding to largest residual. If we continue the iteration, all outliers are eventually removed.

As with all outliers removal processes, there is a risk that discarding a set of outliers will remove some inliers at the same time. In this framework, the outliers are individual pixels, which are in good supply in visual recognition applications. For example, a face image will typically contain hundreds or thousands of pixels. Removing a small fraction of the good pixels is therefore unlikely to affect recognition performance. However, if too many pixels are removed, the remaining pool may be too small for successful recognition. Therefore, we propose a process to restore incorrectly removed pixels where possible, as part of the overall outlier removal algorithm list in Algorithm 1. In practice, the heuristic remedy step does improve the performance of our method on the visual recognition problems in our experiments. Also note that it is impossible that all points in the support set are moved back in step 6, which is because we can prove that

$$\begin{aligned} \max_{i \in I_{supp}} f_i(\beta^*) &\geq \min_{\beta} \max_{i \in I_{supp}} f_i(\beta) \\ &= \min_{\beta} \max_{i \in I_t \cup I_{supp}} f_i(\beta) > \min_{\beta} \max_{i \in I_t} f_i(\beta). \end{aligned}$$

Note that the equality in the last equation can be obtained based on Corollary 1.

5. A fast algorithm for the L_∞ norm minimization problem

Recalling Theorem 2, we may remove any data not in the support set without changing the value of the L_∞ norm. This property allows us to subdivide the large problem into a set of smaller sub-problems. We will proceed by first presenting a useful definition of pseudoconvexity:

Definition 1. A function $f(\cdot)$ is called pseudoconvex if $f(\cdot)$ is differentiable and $\nabla f(\bar{x})(x - \bar{x}) \geq 0$ implies $f(x) \geq f(\bar{x})$.

In this definition, $f(\cdot)$ has to be differentiable. However the notion of pseudoconvexity can be generalized to non-differentiable functions (Boyd, 2001):

Definition 2. A function $f(\cdot)$ is called pseudoconvex if for all $x, y \in \text{dom}(f)$:

$$\begin{aligned} \exists x^* \in \partial f(x) : \quad & \langle x^*, y - x \rangle \geq 0 \\ \Rightarrow \forall z \in [x, y] : & f(z) \leq f(y), \end{aligned} \tag{9}$$

where $\partial f(x)$ is subdifferential of $f(x)$.

Both of these definitions share the property that any local minimum of a pseudoconvex function is also a global minimum. Based on the first definition, it has been proved that if the residual error functions $f_i(\cdot)$ are pseudoconvex and differentiable, the cardinality of the support set is not larger than $d + 1$ (Olsson et al., 2008; Li, 2009). Following the proof in Li (2009), one can easily validate the following corollary:

Corollary 1. For the minimax problem with pseudoconvex residual functions (differentiable or not), there must exist a subset $I_s \subset I, |I_s| \leq d + 1$ such that

$$f_{I_s}(\beta^*) = \min_{\beta} \max_{i \in I_s} f_i(\beta) = f_I(\beta^*) = \min_{\beta} \max_{i \in I} f_i(\beta). \tag{10}$$

It is clear that the squared residual functions $f_i(\beta) = (\mathbf{a}_i \beta - y_i)^2$ in (3) are convex and differentiable, hence also, pseudoconvex. The absolute residual function (5): $f_i(\beta) = |\mathbf{a}_i \beta - y_i|$ is sub-differentiable. It is easy to verify that the only non-differentiable point, the origin, satisfies the second definition. Therefore, the objective function in (5) also satisfies the pseudo-convexity property and hence Corollary 1 holds.

The above corollary says that we can solve a sub-problem with at most $d + 1$ measurements without changing the estimated solution to the original minimax problem. However before solving the sub-problems, we should first determine the support set. We choose to solve a set of small sub-problems using an optimization method called *column generation* (CG). The CG method adds one constraint at a time to the current sub-problem until an optimal solution is obtained.

The process is as follows: We first choose $d + 1$ measurements not contained in the support set. These data are then used to compute a solution and residuals for all the data in the main problem are determined. Then the most violated constraint, or the measurement corresponding to the largest residual, is added to the sub-problem. The sub-problem is then too large, therefore we solve the sub-problem again (now with size $d + 2$) and remove an inactive measurement. Through this strategy, the problem is prevented from growing too large, and violating Corollary 1. When there are no violated constraints, we have obtained the optimal solution.

The proposed fast method is presented in Algorithm 2. We divide the data into an active set, corresponding to a sub-problem, and the remaining set with the L_2 norm minimization. This algorithm allows us to solve the original problem with the measurement matrix of size $n \times d$, by solving a series of small problems with size $(d + 1) \times d$ or $(d + 2) \times d$. In most visual recognition problems, $n \gg d$. Typically the algorithm converges in less than 30 iterations in all of our experiments. We will show that this strategy radically improves computational efficiency.

For maximal efficiency, we choose to solve the LP problem, (6), and utilize the code generator CVXGEN (Matingley and Boyd, 2012) to generate custom, high speed solvers for the sub-problems in Algorithm 2. CVXGEN is a software tool that automatically generates customized C code for LP or QP problems of modest size. CVXGEN is less effective on large problems (e.g., >2000 variables). However, in Algorithm 2 we convert the original problem into a set of small sub-problems, which can be efficiently solved with the generator. CVXGEN embeds the problem size into the generated code, restricting it to fixed-size problems (Matingley and Boyd, 2012). The proposed method only needs to solve problems of size $d + 1$ or $d + 2$, enabling the use of CVXGEN.

6. Experimental results

In this section, we first illustrate the effectiveness and efficiency of our algorithm on several classic geometric model fitting problems. Then the proposed method is evaluated on face recognition problems with both artificial and natural contiguous occlusions. Finally, we test our method on the iris recognition problem, where both segmentation error and occlusions are present. For comparison, we also evaluate several other representative robust regression methods on face recognition problems.

Once the outliers have been removed from the dataset, any solver can be used to obtain the final model estimate. We implemented the original minimax algorithm using Matlab package CVX (Grant and Boyd, 2011) with SeDuMi solver (Sturm, 1999) while the proposed fast algorithm was implemented using solvers generated by CVXGEN (Matingley and Boyd, 2012). All experiments are conducted in Matlab running on a PC with a Quad-Core 3.07 GHz CPU and 12 GB of RAM, using mex to call the solvers from CVXGEN. Note that the algorithm makes no special effort to use multiple cores, though Matlab itself may do so if possible.

Algorithm 2 A fast algorithm for the L_∞ problem

- 1: **Input:** The measurement data matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$; the response vector $\mathbf{y} \in \mathbb{R}^n$; maximum iteration number l_{\max} .
- 2: **Initialization:** Initialize the active set I_s with indices corresponding to the largest $d + 1$ absolute residuals from vector $\{f_i(\beta_{I_s})\}$, $i \in I$ using the LS solution as in (2); set $I_r \leftarrow I \setminus I_s$; set iteration counter $l \leftarrow 0$.
- 3: Solve the L_∞ -minimization sub-problem

$$\beta^* = \arg \min_{\beta} \max_{i \in I_s} f_i(\beta) \quad (11)$$
 and set $t^* \leftarrow \max_{i \in I_s} f_i(\beta^*)$.
- 4: **while** $l < l_{\max}$ **do**
- 5: Get the most violated measurement from the remaining set I_r : $i_m = \arg \max_{i \in I_r} f_i(\beta^*)$
- 6: Check for optimal solution:
if $f_{i_m}(\beta) \leq t^*$, **then** break (problem solved).
- 7: Update the active and remaining set:
 $I_s \leftarrow I_s \cup \{i_m\}$, $I_r \leftarrow I_r \setminus \{i_m\}$.
- 8: Solve the L_∞ -minimization sub-problem in (11).
- 9: Move the inactive measurement index $i_d = \arg \min f_i(\beta^*)$ to the remaining set: $I_s \leftarrow I_s \setminus \{i_d\}$, $I_r \leftarrow I_r \cup \{i_d\}$
- 10: $l \leftarrow l + 1$.
- 11: **end while**
- 12: **Output:** β^* .

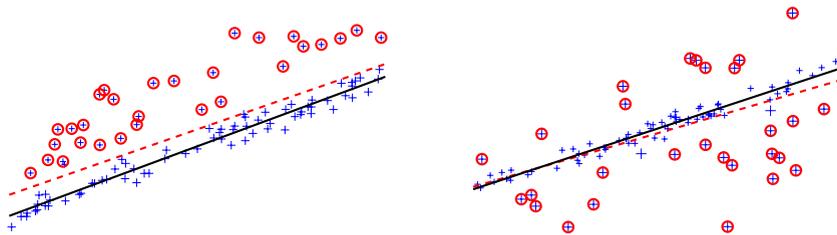


Fig. 1. Two examples demonstrating the performance of our algorithm on data contaminated on one-side (left) and two-sides (right). Outliers detected by our algorithm are marked with circles. The solid black line is the result after outlier removal and the red dashed line shows the result with outliers included. In these two cases, we set $n = 100$ and $k = 70$, where n and k are described in Section 6.1.1.

6.1. Geometric model fitting

6.1.1. Line fitting

Fig. 1 shows estimation performance when our algorithm is used for outlier removal and the line is subsequently estimated via least squares, on data generated under two different error models.

We generate $\mathbf{A} \in \mathbb{R}^{n \times d}$ randomly and $\beta \in \mathbb{R}^d$ randomly. We then set the first k error terms $\epsilon_j, j \in [1, \dots, k]$ as independent standard normal random variables. We set the last $n - k$ error terms $\epsilon_j, j \in [k + 1, \dots, n]$ as independent chi squared random variables with 5 degrees of freedom. We also test using the two-sided contamination model which sets the sign of the last $n - k$ variables randomly such that the outliers lie on both sides of the true regression line. In both cases we set $\mathbf{y} = \mathbf{A}\beta + \epsilon$.

As can be seen in Fig. 1, our method detects all of the outliers and consequently generates a line estimate which fits the inlier set well for both noise models, whilst the estimate obtained with the outliers included achieves a reasonable estimate only for the two-sided contamination case, where the outliers are evenly distributed on both sides of the line.

6.1.2. Ellipse fitting

An example of the performance of our method applied to ellipse fitting is shown in Fig. 2. 100 points were sampled uniformly around the perimeter of an ellipse centred at $(0, 0)$, and were then perturbed via offset drawn from $\mathcal{N}(0, 1)$. 30 outliers were randomly drawn from an approximately uniform distribution within the bounding box shown in Fig. 2. The result of the method, again shown in Fig. 2, shows that our method has correctly identified the inlier and outlier sets, and demonstrates that the centre and radius estimated by our method are accurate.

6.1.3. Efficiency

Next we compare the computational efficiency of the standard L_∞ norm outlier removal process and of the proposed fast algorithm.

For the line fitting problem we generate the data using the scheme described previously. We initially fix the data dimension $d = 2$ and increase the problem size n from 20 to 10 000. The outlier fraction, k , is set 90% of n . A comparison

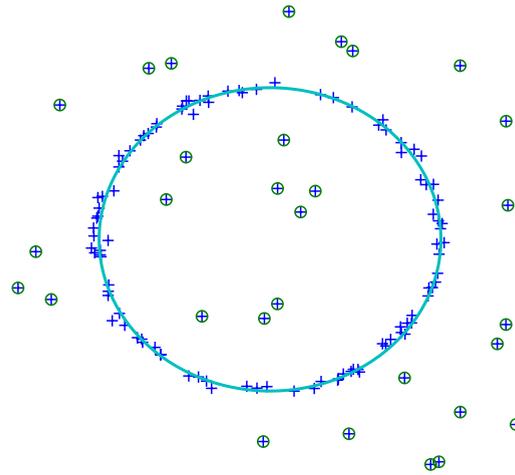


Fig. 2. An example of the performance of our method in ellipse fitting. Points identified as outliers are marked by circles.

Table 1

Computation time (in seconds) comparison of the original and fast algorithms implemented with different solvers. For line fitting problem, we fixed $d = 2$ and the observation number is increased from 20 to 10 000.

Method	Number of observations							
	20	50	100	200	500	1000	2000	10 000
Original	0.185	0.454	0.89	1.926	5.093	11.597	29.968	313.328
Fast	0.002	0.005	0.011	0.031	0.083	0.164	0.395	4.127

Table 2

Computation time (in seconds) comparison of the original and fast algorithms implemented with different solvers. Data is generated as 200 observations with dimension varying from 2 to 10.

Method	Feature dimension				
	2	4	6	8	10
Original	1.926	1.984	1.994	2.016	2.039
Fast	0.031	0.045	0.068	0.096	0.128

of the running time for these two algorithms are shown in Table 1. The fast algorithm finishes 70–80 times faster than the original algorithm. Specifically, with dimension 10 000 the fast algorithm finishes in approximately 4 s, whilst the original algorithm requires more than 5 min. In this case, the proposed fast approach is about 80 times faster than the conventional approach.

Second, we fix the number of observations $n = 200$ and vary the data dimension d from 2 to 10. Execution times are shown in Table 2. Consistent with the last experiment, the fast algorithm completes far more rapidly than the original algorithm in all situations. When $d = 2$, the proposed fast algorithm is faster than the original algorithm by more than 60 times. With a larger dimension $d = 10$, the proposed fast algorithm takes only 0.128 s to complete while the original algorithm requires more than 2 s.

6.2. Robust face recognition

In this section, we test our method on face recognition problems from 3 datasets: AR (Martinez and Benavente, 1998), Extended Yale B (Georghiadis et al., 2001), and CMU-PIE (Sim et al., 2003). A range of state-of-the-art algorithms are compared to the proposed method. Recently, sparse representation based classification (SRC) (Wright et al., 2009) obtained an excellent performance for robust face recognition problems, especially with contiguous occlusions. The SRC problem solves $\min \|\beta\|_1$, s.t. $\|\mathbf{y} - \mathbf{A}\beta\|_2 \leq \epsilon$, where \mathbf{A} is the training data from all classes, β is the corresponding coefficient vector and $\epsilon > 0$ is the error tolerance. To handle occlusions, SRC is extended to $\min \|\omega\|_1$, s.t. $\|\mathbf{y} - \mathbf{B}\omega\|_2 \leq \epsilon$, where $\mathbf{B} = [\mathbf{A}, \mathbf{I}]$, and $\omega = [\beta, \mathbf{e}]$ and \mathbf{e} are the identity matrix and error vector respectively. SRC assigns the test image to the class with smallest residual: $\text{identity}(\mathbf{y}) = \min_i \|\mathbf{y} - \mathbf{A}\delta_i(\beta)\|_2$. Here $\delta_i(\beta)$ is a vector whose only nonzero entries are the entries in β corresponding to the i th class. We also evaluate the following two methods which are related to our method. Most recently, a method called Collaborative Representation-based Classification (CRC) was proposed in Zhang et al. (2011) which relaxes the L_1 norm to L_2 norm. Linear regression classification (LRC) (Naseem et al., 2010) casts face recognition as a simple linear

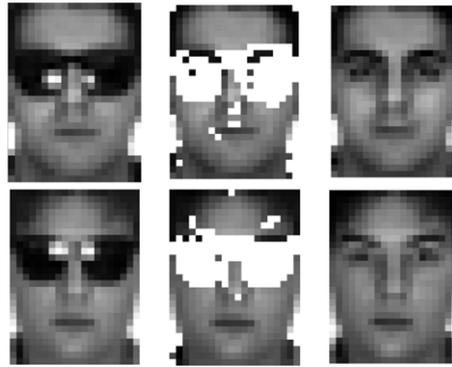


Fig. 3. Detecting the sunglasses occlusion of two example images with dimension 27×20 from the AR dataset. Each row shows an original image of a particular subject, followed by an image where outlying pixels have been automatically marked in white and finally a reconstructed image of the subject.

Table 3

Accuracy rates (%) of different methods on the AR dataset with sunglasses occlusion. The various feature dimensions correspond to downsampling the original 165×120 pixel images to 9×6 , 13×10 , 20×15 and 27×20 , respectively.

Method	Feature dimension			
	54	130	300	540
LRC	21.0%	38.5%	54.5%	60.0%
SRC	48.0%	67.5%	69.5%	64.5%
CRC	22.0%	35.5%	44.5%	56.0%
<i>MM</i> -estimator	0.5%	8.5%	21%	24%
LMS	9%	25%	37.5%	48%
Our method	43.0%	85.0%	99.5%	100%

regression problem: $\min \|\beta_i\|_1$, s.t. $\mathbf{y} = \mathbf{A}_i\beta_i$, where \mathbf{A}_i and β_i are the training data and representative coefficients with respect to class i . LRC selects the class with smallest residual: $\text{identity}(\mathbf{y}) = \min_i \|\mathbf{y} - \mathbf{A}_i\beta_i\|_2$. Both CRC and LRC achieved competitive or even better results than SRC (Zhang et al., 2011; Naseem et al., 2010) in some cases.

For the purposes of the face recognition experiments, outlier pixels are first removed using our method, leaving the remaining inlier set to be processed by any regression based classifier. In the experiments listed below, LRC has been used for this purpose due to its computational efficiency.

Lots of robust regression estimators have been developed in the statistic literature. In this section, we also compare other two popular estimators, namely, Least median of squares (LMS) (Rousseeuw, 1984) and *MM*-estimator (Yohai, 1987), both of which have high-breakdown points and do not need to specify the number of outliers to be removed. Since it is difficult to compute the theoretical LMS and *MM* estimators exactly (especially for visual recognition problems with large data sizes), in practice we use the Forward Search Data Analysis (FSDA) toolbox² to compute the approximations of these two estimators. Comparison is conducted on face recognition problems with both artificial and natural occlusions. These methods are first used to estimate the coefficient β and face images are recognized by the minimal residuals.

6.2.1. Faces recognition despite disguise

The AR dataset (Martinez and Benavente, 1998) consists of over 4000 facial images from 126 subjects (70 men and 56 women). For each subject 26 facial images were taken in two separate sessions, 13 per session. The images exhibit a number of variations including various facial expressions (neutral, smile, anger, and scream), illuminations (left light on, right light on and all side lights on), and occlusion by sunglasses and scarves. Of the 126 subjects available, 100 have been randomly selected for testing (50 males and 50 females) and the images cropped to 165×120 pixels. 8 images of each subject with various facial expressions, but no occlusions, were selected for training. Testing was carried out on 2 images of each of the selected subjects wearing sunglasses. Fig. 3 shows two typical images from the AR dataset with the outliers (30% of all the pixels in the face images) detected by our method set to white. The reconstructed images are shown as the third and sixth images.

The images were downsampled to produce features of 30, 54, 130, and 540 dimensions respectively. Table 3 shows a comparison of the recognition rates of various methods. Our method exhibits superior performance to LRC, CRC and SRC in all except the lowest feature dimension case. Specifically with feature dimension 540, the proposed method achieves a perfect

² <http://ipsc.jrc.ec.europa.eu/?id=732>.

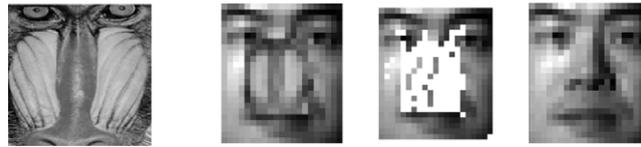


Fig. 4. Detecting the random placed square monkey face (the first image) in an example face image (the second image) from the Extended Yale B dataset. Outliers (30% of the whole image) detected by our method are marked as white pixels (the third image) and the reconstructed image is shown in the rightmost figure.

Table 4

Mean and standard deviations of recognition accuracies (%) in the presence of randomly placed block occlusions of images from the Extended Yale B dataset based on 5 runs' results.

Method	Occlusion rate					
	10%	20%	30%	35%	40%	50%
LRC	98.5±0.18	88.6±0.18	68.8±3.3	60.4±3.2	51.8±1.3	41.7±1.7
SRC	99.0±0.3	94.8±1.2	77.2±1.5	67.3±1.1	56.2±2.2	45.9±1.5
CRC	98.6±0.7	90.3±1.7	74.9±2.6	67.2±1.0	58.3±1.8	48.3±2.0
Our method	99.8±0.1	99.3±0.1	98.2±1.0	96±1.2	91.4±0.7	85.5±0.7

accuracy 100%, which outperforms LRC, CRC and SRC by 40%, 44% and 35.5% respectively. *MM*-estimator and LMS failed to achieve good results on this dataset, which is mainly because the residuals of outliers severely affect the final classification although relatively accurate coefficients could be estimated. In this face recognition application, the final classification is based on the fitting residual. These results highlight the ability of our method for outlier removal, which can significantly improve the face recognition performance.

6.2.2. Contiguous block occlusions

In order to evaluate the performance of the algorithm in the presence of artificial noise and larger occlusions, we now describe testing where large regions of the original image are replaced by pixels from another source. The Extended Yale B dataset (Georghiades et al., 2001) was used as the source of the original images and consists of 2414 frontal face images from 38 subjects under various lighting conditions. The images are cropped and normalized to 192×168 pixels (Lee et al., 2005). Following (Wright et al., 2009), we choose subsets 1 and 2 (715 images) for training and Subset 3 (451 images) for testing. In our experiment, all the images are downsampled to 24×21 pixels. We replace a randomly selected region, covering between 10% and 50% of each image, with a square monkey face. Fig. 4 shows the monkey face, an example of an occluded image, the outlying pixels detected by our method and a reconstructed copy of the input image.

Table 4 compares the average recognition rates of the different methods, averaged over five separate runs. Our proposed method outperforms all other methods in all conditions. With small occlusions, all methods achieve high accuracy, however, the performance of LRC, SRC and CRC deteriorate dramatically as the size of the occlusion increases. In contrast, our method is robust in the presence of outliers. In particular, with 30% occlusion our method obtains 98.2% accuracy while recognition rates of all the other methods are below 80%. With 50% occlusion, all other methods show low performances, while the accuracy for our method is still above 85%. According to Table 4, we can also see that the proposed method is more stable in the sense of accuracy variations, which is mainly because the outliers are effectively detected.

6.2.3. Partial face features on the CMU PIE dataset

As shown in the previous example, occlusion can significantly reduce face recognition performance, particularly in methods without outlier removal. Wright et al. (2009) attempt to identify faces on the basis of particular sub-images, such as the area around the eye or ear, etc. Here, we use the complete face and remove increasing portions of the bottom half of the image, so that initially the neck is obscured, followed by the chin and mouth, etc. The removal occurs by setting the pixels to be black. Thus, the complete image is used as a feature vector, and a subset of the elements is set to zero. A second experiment is performed following the same procedure, but to the central section of the face, thus initially obscuring the nose, then the ears, eyes and mouth, etc.

In this experiment, we use the CMU-PIE dataset (Sim et al., 2003) which contains 68 subjects and a total of 41 368 face images. Each person has his/her picture taken with 13 different poses, 43 different illumination conditions, and with 4 different expressions. In our experiment all of the face images are aligned and cropped, with 256 grey level per pixel (He et al., 2005), and finally resized to 24×24 pixels. Here we use the subset containing images of pose C27 (a nearly front pose) and we use the data from the first 20 subjects, each subject with 21 images. The first 15 images of each subject are used for training and the last 6 images for testing. The test images are preprocessed so that one part (bottom or middle) of faces (from 10% to 40% of pixels) is set to black. See Fig. 5 for examples.

The recognition rates of different methods when the bottom area of the image is occluded are reported in Table 5. When occlusion area is small, all methods except *MM*-estimator obtain perfect 100% recognition rates. When occlusion area increases to 20% of the image size, accuracy for LRC drops to 80%, which is because the black pixels bias the linear regression

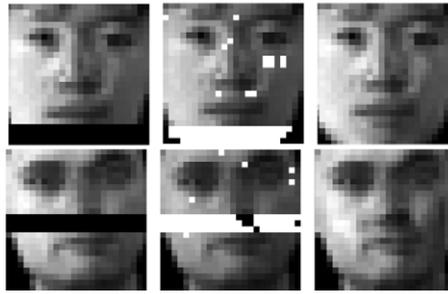


Fig. 5. This figure demonstrates the detection of the dead pixels, covering about 13% of the input image. Each row shows a distinct example drawn from the CMU-PIE dataset. Outliers detected by our method are marked as white pixels in the centre column, followed by the reconstructed images in the last column.

Table 5

Recognition accuracies of various methods on the CMU-PIE dataset with dimension 24×24 . 10%–30% of pixels in the bottom area are replaced with black.

Method	Percentage of image removed		
	10%	20%	30%
LRC	100%	80.0%	61.7%
SRC	100%	98.3%	71.7%
CRC	100%	96.7%	83.3%
<i>MM</i> -estimator	99.2%	41.7%	15.0%
LMS	100%	77.5%	58.3%
Our method	100%	100%	100%

Table 6

Recognition accuracies of various methods on the CMU-PIE dataset with dimension 24×24 . 10%–30% of pixels in the middle area are replaced with black.

Method	Percentage of image removed		
	10%	20%	30%
LRC	100%	91.7%	35.0%
SRC	100%	93.3%	65.0%
CRC	98.3%	87.5%	53.3%
<i>MM</i> -estimator	92.5%	7.5%	0.8%
LMS	100%	80.8%	22.5%
Our method	100%	99.2%	90.0%

estimate. The technique used in SRC mentioned above performs better than LRC when occlusion is present, achieving 98.3% accuracy. However our method is able to achieve 100% accuracy with that level of occlusion. We can see that CRC achieves a relatively good result (83.3%) for 30% occlusion, accuracies of other methods (including robust methods *MM*-estimator and LMS) drop dramatically. In contrast, our method still achieves 100% accuracy which demonstrates the robustness of our method against heavy occlusion. The comparison of these methods for occlusion in the middle part of faces is shown in Table 6. These results again show the robustness of our method against heavy occlusions. Almost all the methods show lower accuracy than in the former situation. Such a result leads to the conclusion that information from the middle part of a face (area around nose) is more discriminative than that from the bottom part (area around chin) for face recognition.

6.2.4. Efficiency

For the problem of identifying outliers in face images, we compare the computation efficiency using the AR face dataset, as described in Section 6.2.1. We vary the feature dimension from 54 to the original 19 800. Table 7 shows the execution time for both the proposed fast algorithm and the original method. We can see that the fast algorithm outperforms the original in all situations. With low dimensional features, below 4800, the fast algorithm is approximately 20 times faster than the original. When the feature dimension increases to 19 800, the original algorithm needs about 1.43 h while the fast algorithm costs only about 6 min.

6.3. Robust iris recognition

Iris recognition is a commonly used non-contact biometric measure used to automatically identify a person. Occlusions can also occur in iris data acquisition, especially in unconstrained conditions, caused by eyelids, eyelashes, segmentation

Table 7

Computation time (in seconds) of the original and fast algorithms when applied to the AR face dataset.

Method	Feature dimension				
	54	300	1200	4800	19 800
Original	2.051	9.894	48.371	396.323	5150.137
Fast	0.113	0.566	2.564	18.811	361.689

Table 8

Classification accuracies on the ND iris dataset. Note that percentage of artificial occlusion 0% means there are only segmentation errors.

Method	Percentage of artificial occlusion		
	0%	10%	20%
LRC	89.5%	43.8%	20.6%
SRC	95.6%	61.3%	51.3%
Our method	96.3%	96.3%	95%

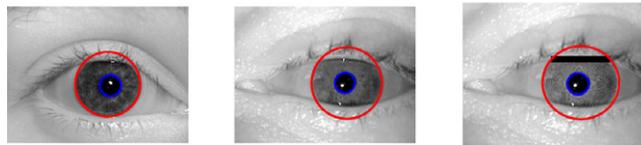


Fig. 6. Three example images from the ND-IRIS-0405 dataset. Features are extracted from the iris area which is between the detected iris and pupil boundaries as shown in red and blue circles respectively. The first two iris images suffer from increasing segmentation errors while the third one suffers from both segmentation error and artificial occlusion. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

errors, etc. In this section we test our method against segmentation errors, which can result in outliers from eyelids or eyelashes. Specially we take the ND-IRIS-0405 dataset (Phillips et al., 2010), which contains 64,980 iris images obtained from 356 subjects with a wide variety of distortions. In our experiment, each iris image is segmented by detecting the pupil and iris boundaries using the open-source package of Masek and Kovesei (2003). 80 subjects were selected and 10 images from each subject were chosen for training and 2 images for testing. To test outlier detection, segmentation errors and artificial occlusions were placed on the iris area, in a similar fashion as (Pillai et al., 2011). A few example images and their detected iris and pupil boundaries are shown in Fig. 6. The feature vector is obtained by warping the circular iris region into a rectangular block by sampling with a radial resolution 20 and angular resolution of 240 respectively. These blocks were then resized to 10×60 . For our method, 10% of pixels are detected and removed when test images are with only segmentation errors, and the corresponding additional number of pixels are removed for artificial occlusions.

The recognition results are summarized in Table 8. SRC used in Pillai et al. (2011) for iris recognition and LRC are compared with our method. We can clearly see that the proposed method achieved the best results with all feature dimensions. Specifically, our method achieves 96.3% accuracy when iris images are with only segmentation errors while accuracy for LRC is 89.5%. SRC performs well (95.6%) for this task. However when 10% additional occlusions occur in the test images, performances for LRC and SRC drop dramatically to 43.8% and 61.3% respectively, while our method still achieves the same result 96.3% as before. When occlusions increase to 20%, our method still obtains a high accuracy 95% which is higher than those of LRC and SRC by 74.4% and 43.7% respectively.

6.3.1. Efficiency

Table 9 shows the computation time comparison of different methods on the iris recognition problem. Consistent with the former results, the proposed algorithm is much more efficient than the original algorithm.

7. Discussion

The main drawback of our method is that one has to first estimate the outlier percentage empirically as done by many other robust regression methods. Actually, to our knowledge, for almost all the *outlier removal* methods, one has to pre-set the outlier percentage or some other parameters such as a residual threshold. This is in contrast with those robust regression methods using a robust loss such as the Huber function or even nonconvex loss. These methods do not need to specify the outlier percentage.

One may be concerned on how the proposed algorithm will perform with an under or over estimated p . To this end, we evaluate our method by varying p from 5% to 45% on three face recognition databases. Tables 10–12 show the results with varying p values on the AR, Extended Yale B and CMU PIE datasets, respectively. We can see that the proposed method is not very sensitive to the pre-estimated outlier percentage when p is over 30%.

Table 9

Computation time (in seconds) comparison of the original and fast algorithms on the Iris dataset with different feature resolutions (shown in the first row).

Method	Image resolution			
	5 × 20	5 × 60	10 × 120	20 × 240
Original	2.400	6.692	31.942	255.182
Fast	0.131	0.385	1.941	14.906

Table 10

Recognition accuracies on the AR dataset with different percentage of outliers removed by our method. The feature dimension is set to 13 × 10 and 20 × 15.

Dimension	Percentage of removed pixels						
	5%	15%	25%	30%	35%	40%	45%
13 × 10	57%	75.5%	82%	85%	92.5%	95%	94.5%
20 × 15	68%	90.5%	98%	99.5%	100%	100%	99.5%

Table 11

Recognition accuracies on the Extended Yale B dataset with different percentage of outliers removed by our method. The images are resized to 24 × 21 pixels and the occlusion rate is set to 30%.

Dimension	Percentage of removed pixels						
	5%	15%	25%	30%	35%	40%	45%
24 × 21	73.5%	89%	97.5%	98.2%	98.2%	98.1%	97.9%

Table 12

Recognition accuracies on the CMU PIE dataset with different percentage of outliers removed by our method. The images are resized to 24 × 24 pixels and 30% of pixels in the bottom area are replaced with black.

Dimension	Percentage of removed pixels						
	5%	15%	25%	30%	35%	40%	45%
24 × 24	71.9%	91.2%	98.3%	100%	100%	99.8%	99.7%

For some applications (especially visual recognition problems), the true number of outliers is difficult to estimate. However, the image dimension is generally large (e.g., more than 100) in visual recognition tasks. In this situation, we suggest that p is set to from 30% to 40%. Removing a small fraction of inliers will not significantly affect the recognition results. In contrast, we can see that the performance degrades when p is set too small (e.g., 5% and 15%), since a large part of occluded pixels are still involved in the estimation. So in practice one should avoid setting p too small.

From Table 10, we also observe that our method becomes more stable when the image resolution is higher. This is mainly because, as mentioned before, visual recognition problems generally supply large amount of pixels by high dimensional images and consequently it is more crucial to reject as many outliers as possible than to keep all inliers.

Different from our approach, there exist many robust estimators which do not need to specify the outliers number, such as *MM*-estimator (Yohai, 1987), *LMS* (Rousseeuw, 1984) and *DPM* (Park et al., 2012). These methods can also be applied to visual recognition problems as we have shown in Section 6. However, the difference is that our method can directly identify the outliers, which can help compute more reliable residuals for classification as shown in Section 6. Of course, for these methods, observations can be detected as outliers when the corresponding standardized residuals exceed the cutoff point, which also has to be determined *a priori* though.

8. Conclusion

In this work, we have proposed an efficient method for minimizing the L_∞ norm based robust least squares fitting, and hence for iteratively removing outliers. The efficiency of the method allows it to be applied to visual recognition problems which would normally be too large for such an approach. The method takes advantage of the nature of the L_∞ norm to break the main problem into more manageable sub-problems, which can then be solved via standard, efficient, techniques.

The efficiency of the technique and the benefits that outlier removal can bring to visual recognition problems were highlighted in the experiments, with the computational efficiency and accuracy of the resultant recognition process easily beating all other tested methods.

Like many other robust fitting methods, the proposed method needs a parameter: the number of outliers to be removed. One may heuristically determine this value. Although it is not very sensitive for the visual recognition problems, in the future, we plan to investigate how to automatically estimate the outlier rate in the noisy data.

Acknowledgements

This work was in part supported by ARC Future Fellowship FT120100969. F. Shen's contribution was made when he was visiting The University of Adelaide.

References

- Basri, R., Jacobs, D., 2003. Lambertian reflectance and linear subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 218–233.
- Boyd, S., Vandenberghe, L., 2004. *Convex Optimization*. Cambridge University Press.
- Fidler, S., Skocaj, D., Leonardis, A., 2006. Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (3), 337–350.
- Fidler, S., Skocaj, D., Leonardis, A., 2006. Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (3), 337–350.
- Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 381–395.
- Floudas, C., Pardalos, P. (Eds.), 2001. Generalized monotone multivalued maps. In: *Encyclopedia of Optimization*, pp. 764–769.
- Georgiades, A.S., Belhumeur, P.N., Kriegman, D.J., 2001. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (6), 643–660.
- Grant, M., Boyd, S., CVX: Matlab software for disciplined convex programming, version 1.21, <http://cvxr.com/cvx>, 2011.
- Hartley, R.I., Schaffalitzky, F., 2004. L_∞ minimization in geometric reconstruction problems. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*.
- He, X., Cai, D., Niyogi, P., 2005. Laplacian score for feature selection. In: *Proc. Advances Neural Info. Process. Syst.*
- Huber, P.J., 1973. Robust regression: asymptotics, conjectures and monte carlo. *Ann. Math. Statist.* 1 (5), 799–821.
- Huber, P.J., 1981. *Robust Statistics*. Wiley, New York, p. 2005.
- Hubert, M., Rousseeuw, P.J., Aelst, S.v., 2008. High-breakdown robust multivariate methods. *Statist. Sci.* 23 (1), 92–119.
- Jureckova, J., 1971. Nonparametric estimate of regression coefficients. *Ann. Math. Statist.* 42 (4), 1328–1338.
- Kahl, F., 2005. Multiple view geometry and the L_∞ -norm. In: *Proc. Int. Conf. Computer Vision*, Vol. 2, pp. 1002–1009.
- Ke, Q., Kanade, T., 2005. Quasiconvex optimization for robust geometric reconstruction. In: *Proc. IEEE Conf. Computer Vision*, Vol. 2, pp. 986–993.
- Koenker, R., Portnoy, S., 1987. L -estimation for linear models. *J. Amer. Statist. Assoc.* 82, 851–857.
- Lee, K., Ho, J., Kriegman, D., 2005. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (5), 684–698.
- Li, H., 2009. Efficient reduction of l_∞ geometry problems. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2695–2702.
- Li, H., 2007. A practical algorithm for l_∞ triangulation with outliers. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*.
- Lübbecke, M.E., Desrosiers, J., 2005. Selected topics in column generation. *Oper. Res.* 53 (6), 1007–1023.
- Martinez, A.M., Benavente, R., 1998. *The AR Face Database CVC*, Tech. Rep.
- Masek, L., Kovsi, P., MATLAB source code for a biometric identification system based on iris patterns, 2003.
- Mattingley, J., Boyd, S., 2012. CVXGEN: a code generator for embedded convex optimization. *Optim. Eng.* 13, 1–27.
- Naseem, I., Togneri, R., Bennamoun, M., 2010. Linear regression for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (11), 2106–2112.
- Naseem, I., Togneri, R., Bennamoun, M., 2012. Robust regression for face recognition. *Pattern Recognit.* 45 (1), 104–118.
- Olsson, C., Enqvist, O., Kahl, F., 2008. A polynomial-time bound for matching and registration with outliers. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*.
- Olsson, C., Eriksson, A., Hartley, R., 2010. Outlier removal using duality. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1450–1457.
- Olsson, C., Eriksson, A., Kahl, F., 2007. Efficient optimization for L_∞ -problems using pseudoconvexity. In: *Proc. Int. Conf. Computer Vision*.
- Park, Y., Kim, D., Kim, S., 2012. Robust regression using data partitioning and M -estimation. *Commun. Stat. - Simul. Comput.* 41 (8), 1282–1300.
- Phillips, P.J., Scruggs, W.T., O'Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., Sharpe, M., 2010. FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 831–846.
- Pillai, J., Patel, V., Chellappa, R., Ratha, N., 2011. Secure and robust iris recognition using random projections and sparse representations. *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (9), 1877–1893.
- Rousseeuw, P.J., 2014. Least median of squares regression. *J. Amer. Statist. Assoc.* 79.
- Rousseeuw, P.J., Leroy, A.M., 1987. *Robust regression and outlier detection*. Wiley & Sons, Inc., New York, NY, USA.
- Shi, Q., Eriksson, A., van den Hengel, A., Shen, C., 2011. Is face recognition really a compressive sensing problem? In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 553–560.
- Sim, T., Baker, S., Bsat, M., 2003. The CMU pose, illumination, and expression database. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 1615–1618.
- Sim, K., Hartley, R., 2006. Removing outliers using the L_∞ norm. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 485–494.
- Sturm, J.F., 1999. Using sedumi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.* 11–12, 625–653.
- Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y., 2009. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 210–227.
- Yang, J., Zhang, L., Xu, Y., Yang, J., 2012. Beyond sparsity: the role of l_1 -optimizer in pattern classification. *Pattern Recognit.* 45 (3), 1104–1118.
- Yohai, V.J., 1987. High breakdown-point and high efficiency robust estimates for regression. *Ann. Stat.* 642–656.
- Zhang, L., Yang, M., Feng, X., 2011. Sparse representation or collaborative representation: which helps face recognition? In: *Proc. Int. Conf. Computer Vision*, pp. 471–478.