



ELSEVIER

Available at

[www.ElsevierComputerScience.com](http://www.ElsevierComputerScience.com)

POWERED BY SCIENCE @ DIRECT®

Pattern Recognition Letters 24 (2003) 2953–2966

---

---

Pattern Recognition  
Letters

---

---

[www.elsevier.com/locate/patrec](http://www.elsevier.com/locate/patrec)

# Using symmetry in robust model fitting

Hanzi Wang <sup>\*</sup>, David Suter

*Department of Electrical and Computer Systems Engineering, Monash University, Wellington Road, Clayton, Vic. 3800, Australia*

Received 27 August 2002; received in revised form 15 April 2003

---

## Abstract

The pattern recognition and computer vision communities often employ robust methods for model fitting. In particular, high breakdown-point methods such as least median of squares (LMedS) and least trimmed squares (LTS) have often been used in situations where the data are contaminated with outliers. However, though the breakdown point of these methods can be as high as 50% (they can be robust to up to 50% contamination), they can break down at unexpectedly lower percentages when the outliers are clustered. In this paper, we demonstrate the fragility of LMedS and LTS and analyze the reasons that cause the fragility of these methods in the situation when a large percentage of clustered outliers exist in the data. We adapt the concept of “symmetry distance” to formulate an improved regression method, called the least trimmed symmetry distance (LTSD). Experimental results are presented to show that the LTSD performs better than LMedS and LTS under a large percentage of clustered outliers and large standard variance of inliers.

© 2003 Elsevier B.V. All rights reserved.

*Keywords:* Robust regression; Symmetry distance; Clustered outliers; Breakdown point

---

## 1. Introduction

One major task of computer vision, pattern recognition, machine learning, and related areas: is to fit a model to noisy data (with outliers) (Fischler and Bolles, 1981; Rousseeuw, 1984; Rousseeuw and Leroy, 1987; Zhang, 1997; Danuser and Stricker, 1998; Stewart, 1999). It is common to employ “regression analysis” to undertake such

tasks (Rousseeuw and Leroy, 1987). The most common form of regression analysis is the least squares (LS) method, which can achieve optimum results under Gaussian distributed noise. But this method is extremely sensitive to outliers (gross errors or samples belonging to another structure and distribution). The breakdown point of an estimator may be roughly defined as the smallest percentage of outlier contamination that can cause the estimator to produce arbitrarily large values. Mathematically, let  $Z$  be any sample of  $n$  data points  $(x_1, y_1), \dots, (x_n, y_n)$ ,  $Z = \{z_1, \dots, z_n\}$  and  $z_i = \{x_i, y_i\}$ . For  $m \leq n$ , the finite-sample breakdown point of a regression estimator  $T$  can be written as (Rousseeuw and Leroy, 1987, pp. 10):

---

<sup>\*</sup> Corresponding author. Tel.: +61-395-159471; fax: +61-399-053454.

*E-mail addresses:* [hanzi.wang@eng.monash.edu.au](mailto:hanzi.wang@eng.monash.edu.au) (H. Wang), [d.suter@eng.monash.edu.au](mailto:d.suter@eng.monash.edu.au) (D. Suter).

$$\varepsilon_n^*(T, Z) = \min \left\{ \frac{m}{n}; \sup_{Z^* \in Z_m} \|T_n(Z^*)\| = \infty \right\} \quad (1.1)$$

Because one single outlier is sufficient to force the LS estimator to produce arbitrarily large value, the LS estimator has a breakdown point of 0%.

Since data contamination is usually unavoidable (due to faulty feature extraction, sensor noise and failure, segmentation errors, etc.), there has recently been a general recognition that algorithms should be robust (Haralick, 1986). Robust regression methods are a class of techniques that can tolerate gross errors (outliers). Some robust methods also have a high breakdown point.

Several robust estimators with apparently high breakdown points have been developed during the past three decades. The least median of squares (LMedS) estimator and the least trimmed squares (LTS) estimator (Rousseeuw, 1984; Rousseeuw and Leroy, 1987) are the two most popular methods with claimed high breakdown points, and are currently used most widely in computer vision field (Roth and Levine, 1990; Bab-Hadiashar and Suter, 1998; Ming and Haralick, 2000, etc.). They are based on the idea that the correct fit will correspond to the one with the least median of residuals (for LMedS), or the least sum of trimmed squared residuals (for LTS). The essence of the argument claiming a high breakdown point for the LMedS is that if the uncontaminated data are in the majority, then the median of the residuals should be unaffected by the outliers, and thus the median residual should be a reliable measure of the quality of the fit. Likewise, since the LTS method relies only on (the sum of squares of) the  $h$  smallest residuals, for some choice of the parameter  $h$ , it is thought that this should be robust to contamination so long as  $h$  data points, at least, belong to the true fit. Though the robustness of these methods, compared with patently non-robust methods such as Least Squares, is now widely demonstrated and acknowledged, it is perhaps less widely recognized that LMedS and LTS can breakdown at surprisingly low contamination if those outliers are clustered. Due to the affects of clustered outliers, the correct fit may not correspond to the fit with the least median of squared residual (for LMedS) or the least trimmed squared

residuals (for LTS). It is worth mentioning that this phenomenon is not limited to the LMedS, and LTS. It also happens to most other robust estimators such as random sample consensus—RANSAC (Fischler and Bolles, 1981), residual consensus estimator—RESC (Yu et al., 1994), adaptive least  $k$ th estimator—ALKS (Lee et al., 1998), etc. The mechanism of the breakdown in these robust estimators is similar to that of the LMedS and LTS (see Wang and Suter, 2003). In this paper, we restrict ourselves to the better known and more widely applied LMedS and LTS. Making clear the reasons that cause the breakdown of the LMedS and LTS will help understand the fragility of other robust estimators to clustered outliers.

The key to salvaging the robustness of LMedS and LTS, even in the presence of clustered outliers, can be recognize that LMedS and LTS (and some other robust estimators such as M-estimators, ALKS, etc.) only depend upon a single statistical property of the data (the median of the squared residuals, or the sum of the smallest  $h$  residuals, respectively). If one expects some other property of the data, associated with the true fit, to hold, then one may seek to incorporate a measure of that property into the problem formulation. In this paper, we restrict ourselves to one such property—symmetry.

Symmetry is very common and important in our world. When we fit circles, ellipses, or any symmetric object, one of the most basic features in the model is symmetry. In our method, we introduce symmetry into the model fitting and thereby propose an improved method—the least trimmed symmetry distance (LTSD). The LTSD is influenced not only by the sizes of the residuals of data points, but also by the symmetry of the data points and has applications where one is trying to fit a symmetric model (e.g. circle and ellipses). Experiments show that the LTSD method works better than LMedS and LTS under a large number of clustered outliers.

The main contributions of this paper are as follows:

1. We illustrate situations where LMedS and LTS fail to correctly fit the data in the presence of

clustered outliers, and analyze the reasons that cause the breakdown of these two methods. This provides an important cautionary note when employing these two robust estimators in situations where the outliers are clustered.

2. We introduce the concept of symmetry distance (SD) into model fitting. The concept of SD in computer vision is not novel. However it is a novel concept in the field of model fitting. Based on Su et al.'s *point SD* (2001), we propose a novel symmetry distance and apply it to model fitting.
3. We experimentally show that the proposed method work better than LMedS and LTS under a large percentage of clustered outliers for both simulated and real data.

This paper is organized as follows: in Section 2, several robust estimators are reviewed. The reasons why both LMedS and LTS fail to fit a model under a large percentage of clustered outliers are explored. In Section 3 a novel SD measure is given and our proposed method is developed in Section 4. Experiments demonstrating the utility of the approach (for circle fitting and ellipses fitting) are given in Section 5. Finally, the contributions of this paper and future work are summarized in Section 6.

## 2. Robust estimators

The classical linear model can be written in the form:

$$y_i = x_{i1}\theta_1 + \dots + x_{ip}\theta_p + e_i (i = 1, \dots, n) \quad (2.1)$$

where the error term,  $e_i$ , is usually assumed to be normally distributed with mean zero and standard deviation. The aim of regression analysis is to estimate  $\theta = (\theta_1, \dots, \theta_p)^t$  from the data  $(x_{i1}, \dots, x_{ip}, y_i)$ .

The ordinary LS method estimates  $\hat{\theta}$  by

$$\min_{\hat{\theta}} \sum_{i=1}^n r_i^2 \quad (2.2)$$

where the residual of the  $i$ th datum:  $r_i = y_i - x_{i1}\hat{\theta}_1 - \dots - x_{ip}\hat{\theta}_p$ .

Although the LS method has low computational cost and high efficiency when the data are Gaussian distributional, it is very sensitive to outliers. In order to reduce the influence of outliers, a number of robust estimators have been developed. Among these estimators, the maximum-likelihood-type estimators, i.e. the M-estimators, are best known (Huber, 1973, 1981).

### 2.1. M-estimators

The essence of the M-estimator is to replace the squared residuals  $r_i^2$  in (2.2) by another function of the residuals:

$$\min_{\hat{\theta}} \sum_{i=1}^n \rho(r_i) \quad (2.3)$$

where  $\rho(\cdot)$  is a symmetric, positive-definite function with a unique minimum at zero. Different choices of  $\rho(r_i)$  will yield different M estimators. Unfortunately, it has been proved that M-estimators have a breakdown point of at most  $1/(p + 1)$ , where  $p$  represents the dimension of parameter vector (Li, 1985). This means that the breakdown point will diminish when the dimension of the parameter vector increases.

### 2.2. The repeated median method

Before the repeated median (RM) estimator, it was controversial whether it was possible to find a robust estimator with a high breakdown point. In 1982, Siegel proposed the RM estimator with a high breakdown point of 50% (Siegel, 1982).

The RM method can be summarized as followed: take any  $p$  observations ( $p$  is the dimension of the model),  $(x_{i1}, y_{i1}), \dots, (x_{ip}, y_{ip})$ . The parameter vector  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_p)^t$  can be calculated from this set of data points. The  $j$ th coordinate of this vector is denoted by  $\theta_j(i_1, \dots, i_p)$ . Then the RM estimator is defined as:

$$\hat{\theta} = \text{med}_{i_1}(\dots(\text{med}_{i_{p-1}}(\text{med}_{i_p}(\theta_j(i_1, \dots, i_p)))) \dots) \quad (2.4)$$

The RM estimator has succeeded in solving the problems with small  $p$ . But the time complexity of the RM estimator is  $O(n^p \log^p n)$ , which is very high for multidimensional models and so the

utility of the method is restricted to low dimensional models. Since this method is not widely used in the computer vision community, we do not consider it further in this paper.

### 2.3. The least median of squares method

Rousseeuw proposed the LMedS in 1984. The LMedS method has excellent global robustness and claimed high breakdown point (also 50%). Over the last two decades, LMedS has been growing in popularity. For example, Kumar and Hanson (1989) used the LMedS to solve the pose estimation problem; Roth and Levine (1990) employed it for range image segmentation; Meer et al. (1990) applied it for image structure analysis in the piecewise polynomial field; Zhang (1997) used the least median squares in conic fitting; and Bab-Hadiashar and Suter (1998) employed it for optic flow calculation.

The LMedS method is based on the simple idea of replacing the sum in the least sum of squares formulation by a median. That is, the LMedS estimate is given by

$$\min_{\theta} \operatorname{med}_i r_i^2 \quad (2.5)$$

A drawback of the LMedS method is that it estimates the parameters by solving this non-linear minimization problem. No explicit formula exists for the solution of this problem—the exact solution can only be determined by a search in the space of all possible estimates. There are  $O(n^p)p$ -tuples of data, and it takes  $O(n \log n)$  time to find the median of the residuals of the whole data for each  $p$ -tuple. Thus it costs  $O(n^{p+1} \log n)$  for the LMedS method. In order to reduce the cost to a feasible value, a Monte Carlo type technique, as described next, is usually employed.

A  $p$ -tuple is “clean” if it consists of  $p$  good observations without contamination by outliers. One performs,  $m$  times, random selections of  $p$ -tuples, where one chooses  $m$  so that the probability ( $P$ ) that at least one of the  $m$   $p$ -tuples is “clean” is almost 1. Let  $\varepsilon$  be the fraction of outliers contained in the whole set of points. The probability  $P$  can be expressed as follows:

$$P = 1 - (1 - (1 - \varepsilon)^p)^m \quad (2.6)$$

Thus one can determine  $m$  for given values of  $\varepsilon$ ,  $p$  and  $P$  by:

$$m = \frac{\log(1 - P)}{\log[1 - (1 - \varepsilon)^p]} \quad (2.7)$$

The relative efficiency of the LMedS method is poor when Gaussian noise is present in the data. The LMedS has a low convergence rate of order  $n^{-1/3}$ . Rousseeuw improved the LMedS method by further carrying out a weighted LS procedure after the initial LMedS. In this adaptation, a preliminary scale estimate is found by:

$$S = 1.4826(1 + 5/(n - p))M_i \quad (2.8)$$

where  $M_i$  is the median of the residuals returned by the LMedS procedure.

The weight function  $W_i$  which will be assigned to the  $i$ th data point is usually given by:

$$W_i = \begin{cases} 1 & r_i^2 \leq (2.5S)^2 \\ 0 & r_i^2 > (2.5S)^2 \end{cases} \quad (2.9)$$

The data points corresponding to  $W_i = 0$  are likely to be outliers. The data points having  $W_i = 1$  are inliers and will influence the weighted LS estimate:

$$\min_{\theta} \sum_{i=1}^n W_i r_i^2 \quad (2.10)$$

The LMedS method may be locally unstable when fitting models to data. This means that an infinitesimal change in the data can greatly alter the output (Thomas and Simon, 1992).

### 2.4. The least trimmed squares method

The LTS method was introduced by Rousseeuw (1984), Rousseeuw and Leroy (1987) to improve the low efficiency of LMedS. The LTS estimator can be expressed as

$$\min_{\theta} \sum_{i=1}^h (r^2)_{i:n} \quad (2.11)$$

where  $(r^2)_{1:n} \leq \dots \leq (r^2)_{n:n}$  are the ordered squared residuals,  $h$  is the trimming constant.

The method uses  $h$  data points (out of  $n$ ) to estimate the parameters. The coverage value,  $h$ , may be set from  $n/2$  to  $n$ . The aim of LTS estimator is to find the  $h$ -subset with smallest LS

residuals and use the  $h$ -subset to estimate parameters of models. The claimed LTS breakdown point is  $(n - h)/n$ . When  $h$  is set to  $n/2$ , the LTS estimator has a claimed high breakdown value of 50%.

The advantages of LTS over LMedS are:

- It is less sensitive to local effects than LMedS, i.e. it has more local stability.
- LTS has better statistical efficiency than LMedS (Hössjer, 1994). It converges like  $n^{-1/2}$ .

Due to these merits, LTS is preferred to LMedS (Rousseeuw and Aelst, 1999; Ming and Haralick, 2000).

LTS is also implemented by using random sampling, because the number of all possible  $h$ -subsets ( $C_n^h$ ) grows fast with  $n$ . There are two commonly employed ways to generate a  $h$ -subset (Rousseeuw and Driessen, 1999):

1. Directly generate a random  $h$ -subset from the data points  $n$ .
2. Firstly generate a random  $p$ -subset. If the rank of this  $p$ -subset is less than  $p$ , randomly add data points until the rank is equal to  $p$ . Next, use this subset to compute parameters  $\hat{\theta}_j (j = 1, \dots, p)$  and residuals  $r_i (i = 1, \dots, n)$ . Sort the residuals into  $|r(\pi(1))| \leq \dots \leq |r(\pi(h))| \leq \dots \leq |r(\pi(n))|$ , and  $h$ -subset is set to:  $H := \{\pi(1), \dots, \pi(h)\}$ .

Although the first way is easier than the second, the  $h$ -subset yielded by the first method may contain a lot of outliers. Indeed, the chance of generating a “clean”  $h$ -subset by method (1) tends to zero with increasing  $n$ . In contrast, it is easier to find a “clean”  $p$ -subset without outliers by method (2). Therefore, method (2) can generate more (good) initial subsets with size  $h$  than method (1).

Like LMedS, the efficiency of LTS can be improved by adopting a weighted LS refinement as the last stage.

### 2.5. Factors affecting the achievements of LMedS and LTS

The LMedS method and the LTS method are based on the idea that the correct fit is determined

by a simple situation: the least median of the squared residuals (for LMedS), or by the least sum of trimmed squared residuals (for LTS); and that such a statistic is not influenced by the outliers.

Consider the contaminated distributions as follows (Hampel et al., 1986; Haralick, 1986):

$$F = (1 - \varepsilon)F_0 + \varepsilon H \quad (2.12)$$

where  $F_0$  is an inlier distribution, and  $H$  is an outlier distribution.

The equation above is also called the *gross error model*. When the standard variance of  $F_0$  is small ( $\ll 1$ ) and that of  $H$  is large or  $H$  is uniform distributed, the assumptions leading to the robustness of LMedS or LTS, are true. However, when  $F_0$  is “scattered”, i.e. the standard variance of  $F_0$  is big, and  $H$  is clustered distributed with high density, the assumption is not always true.

Let us investigate an example. In Fig. 1, 100 good data (inliers) with noise distribution  $F_0$  (bivariate normal with unit variance) were generated by adding the noise to samples of a circle with radius 10.0 and center at (0.0, 0.0). Then 80 clustered outliers were added, possessing a spherical bivariate normal distribution with one unit standard variance and mean (20.0, 6.0). As Fig. 1 shows, both LMedS and LTS failed to fit the circle: LMS returned the result with a radius equal to 7.4239 and the center was located at (13.1294,

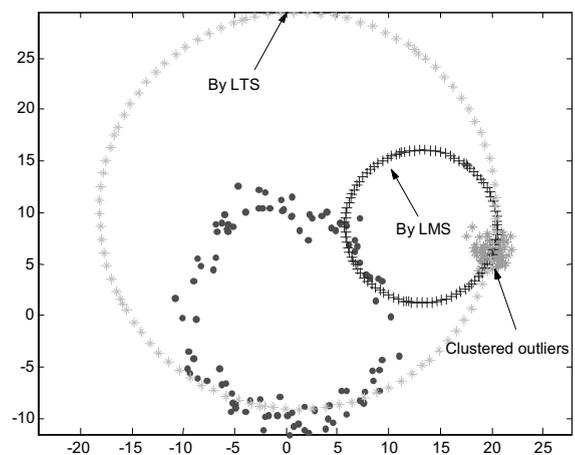


Fig. 1. An example where LMedS and LTS ( $h$  is set to  $0.5n$ ) fail to fit a circle yet there are under 44.5% outliers—though, of course, the outliers are clustered.

8.6289). The results obtained by LTS were: the radius equalled 19.3069 and the center was at (1.1445, 10.1470).

It is important to point out that the failure is inherent, and not simply an artefact of our implementation. Let us check the median of the residuals (for LMedS) and the sum of trimmed squared residuals (for LTS) and we will understand why LMedS and LTS failed to fit to the circle. The median of residuals of the perfect fit is 5.7928. However the median of residuals of final result by the LMedS method is 5.1479.

In fact, during the searching procedure, the LMedS estimator consistently minimises the median of the residuals, starting with initial fits that have a larger median residual than the true fit, but successively finding fits with lower median residuals—proceeding to even lower median residuals than that possessed by the true fit (shown in Fig. 2).

The reason that LTS failed is similar. LTS finds the fit with smallest trimmed squared residuals. The value of the least sum of trimmed squared

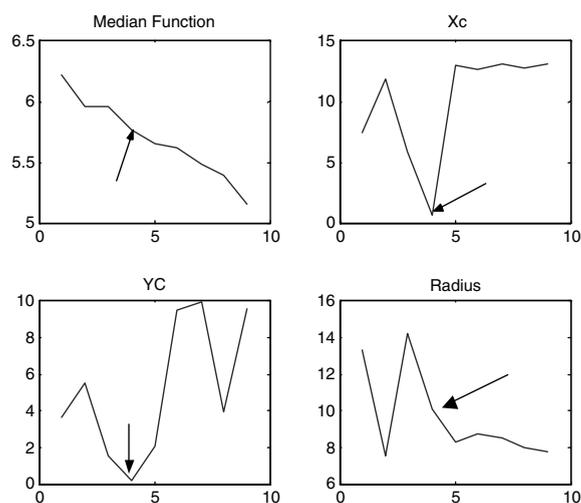


Fig. 2. These graphs show the time evolution of the parameters: median of residuals (top left), centre of the fitted circle ( $x$  coordinate and  $y$  coordinate in the top right and bottom left, respectively), and the radius of the fitted circle (bottom right) as LMedS searches for the “best” fit with the least median of residuals. (Note: the iterations pass by the correct fit (pointed out by arrows) proceeding to fits with even lower median of residuals.)

residuals obtained is 32.6378. However, the same statistic for the “true fit” is 58.6688. Clearly, LTS has “correctly”, by its criterion, obtained a “better” fit (but in fact, the wrong one). The problem is not with the implementation but with the criterion.

Now, let us consider another example showing that the results of the LMedS and the LTS are affected by the standard variance of the inliers. We generated a circle with radius 10.0 and center at (0.0, 0.0). In addition, clustered outliers were added to the circle with mean (20.0, 6.0) and unit standard variance. In total, 100 data points were generated. At first, we assigned 100 data to the circle without any outliers. Then we repeatedly moved two points from the circle to the clustered outliers until 50 data were left in the circle. Thus, the percentage of outliers changed from 0% to 50%. In addition, for each percentage of clustered outliers, we varied the standard variance of the inliers from 0.4 to 1.6 with a step size of 0.3. Fig. 3(a) illustrates one example of the distribution of the data, with 38% clustered outliers and the standard variance of inliers 1.3.

From Fig. 3(b) and (c), we can see that when the standard variance of inliers is no more than 1.0, LMedS can give the right results under high percentage of outliers (more than 44%). However, when the standard variance of inliers is more than 1.0, LMedS does not give the right result even when the percentage of outliers is less than 40%. From Fig. 3(b), we can see when the standard variance of inliers is 0.4, the LTS estimator can correctly give the results even under 50% clustered outliers; while when the standard variance of inliers is 1.6, LTS does not give the right results even when only 30% of the data are outliers.

From the discussion above, we now see several conditions under which LMedS and LTS failed to be robust. A crucial point is: these methods measure only one single statistic: the least median of residuals or the least sum of trimmed squared of residuals, omitting other characteristics of the data. If we look at the failures, we can see that the results lost the most basic and common feature of the inliers with respect to the fitted circle—symmetry.

Symmetry is considered a pre-attentive feature that enhances recognition and reconstruction of

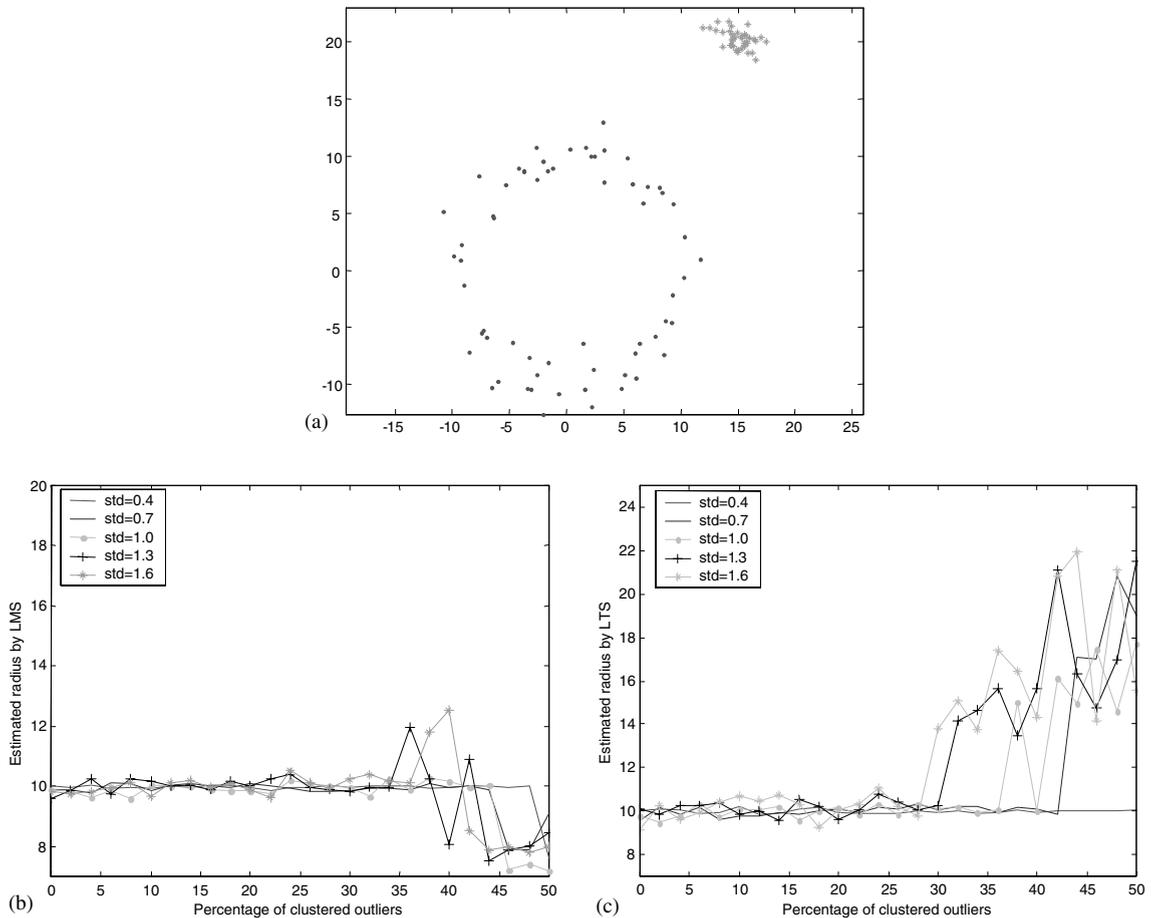


Fig. 3. One case of the distribution of data (a); the results of LMedS (b) and LTS (c) in circle fitting will be affected by the standard variance of inliers and percentages of clustered outliers.

shapes and objects (Attneave, 1995). Symmetry exists in many man-made and natural objects. In the next section, we will introduce the concept of SD into robust regression methods and propose an improved method, called the LTSD, by which the better performance is acquired even when data include clustered outliers.

### 3. The symmetry distance

Symmetry exists almost at all places of the world. A square, a cube, a sphere, and a lot of geometric patterns show symmetry. Architecture usually displays symmetry. Symmetry is also an

important parameter in physical and chemical processes and is an important criterion in medical diagnosis. Even we human beings show symmetry, (for instance, our faces and bodies are roughly symmetrical between right and left). One of the most basic features in the shapes of models we often fit/impose on our data, e.g. circles and ellipsis, is the symmetry of the model. Symmetric data should suggest symmetric models and data that is symmetrically distributed should be preferred as the inlier data (as opposed to the outliers). For decades, symmetry has widely been studied in computer vision community (Bigun, 1988; Marola, 1989; Nalwa, 1989; Kirby and Sirovich, 1990; Reisfeld et al., 1992; Zabrodsky

et al., 1995; Su and Chou, 2001). More detailed definitions of symmetry can be found in (Zabrodsky, 1993). We demonstrate here that symmetry can also be used as a feature to enhance the performance of robust estimators when fitting models with symmetric structure.

### 3.1. The symmetry distance

The exact mathematical definition of symmetry (Miller, 1972; Weyl, 1952) is insufficient to describe and quantify symmetry found both in the natural world and in the visual world.

Su and Chou (2001) proposed a SD measure based on the concept of “point symmetry”. Given  $n$  points  $x_i, i = 1, \dots, N$  and a reference vector  $C$  (e.g. the centroid of the data), the point SD between a point  $x_j$  and  $C$  is defined as follows:

$$d_s(x_j, C) = \min_{\substack{i=1, \dots, N \\ \text{and } i \neq j}} \frac{\|(x_j - C) + (x_i - C)\|}{\|x_j - C\| + \|x_i - C\|} \quad (3.1)$$

From (3.1) we can see that the point SD is non-negative definition. In essence, the measure tries to “balance” data points with others symmetric about the centroid—for example,  $x_i = (2C - x_j)$  exists in the data,  $d_s(x_j, C) = 0$ .

However, according to (3.1), one point could be used repeatedly as the “balancing point” with respect to the center. This does not seem to properly capture the notion of symmetry.

In order to avoid one point being used as a “symmetric point” more than one time by other points, we refine the point SD between a point  $x_j$  and  $C$  as follows:

$$D_s(x_j, C) = \min_{\substack{i=1, \dots, N \\ \text{and } i \neq j \\ \text{and } i \notin \mathfrak{R}}} \frac{\|(x_j - C) + (x_i - C)\|}{\|x_j - C\| + \|x_i - C\|} \quad (3.2)$$

where  $\mathfrak{R}$  is a set of points that have been used as “symmetric point”.

Based on the concept of “point SD”, we propose a non-metric SD. Given a pattern  $x$  consisted of  $n$  points  $x_1, \dots, x_n$  and a reference vector  $C$ , the symmetry distance of the pattern  $x$  with respect to the reference vector  $C$  is:

$$SD_n(x, C) = \frac{1}{n} \sum_{i=1}^n D_s(x_i, C) \quad (3.3)$$

When the SD of a pattern is equal to 0.0, the pattern is perfect symmetric; when the SD of a pattern is very big, the pattern has little symmetry.

## 4. The proposed method

We proposed a new method, which couples the LTS method with the symmetry distance measure. Besides residuals, we also choose SD as a criterion in the model fitting. For simplicity, we call the proposed method the LTSD (least trimmed SD). Mathematically, the LTSD estimate can be written as:

$$\hat{\theta} = \arg \min_{\theta, C} SD_h(x, C) \quad (4.1)$$

Only  $h$  data points with the smallest sorted residuals are used to calculate the SD. The estimated parameters correspond to the least SD.

The specific procedures of the proposed method is given as follows:

- Step 1. Set repeat times (RT) according to equation (2.7). Initialise  $h$  with  $[(n + p + 1) / 2] \leq h \leq n$ . If we want LTSD to have a high breakdown point, say 50%, we can set  $h = (n + p + 1) / 2$ .
- Step 2. Randomly choose  $p$ -subsets, and extend to a  $h$ -subset  $H_1$  by the method (2) in Section 2.4.
- Step 3. Compute  $\hat{\theta}_1$  by LS method based on  $H_1$ . Compute symmetry distance  $SD_1$  based on  $\hat{\theta}_1$  and  $H_1$  using (3.3) in Section 3 and using the centre of the fit (circle or ellipse) as the reference vector  $C$ . Decrement RT and if RT is smaller than 0, go to step 4, otherwise, go to step 2. We calculate the parameters  $\hat{\theta}$  based on  $h$ -subset instead of  $p$ -subset in order to improve the statistical efficiency.
- Step 4. Finally, output  $\hat{\theta}$  with the lowest SD.

### 5. Experimental results

In this section, we will show several examples using the proposed method to fit a model with symmetrical structures. Circle fitting and ellipse fitting have been very popular topics in the computer vision field. One of the obvious characteristics of circles and ellipses is that they are symmetric. We first present an example of circle fitting, to provide insights into the proposed method. We then present a relatively more complicated example of ellipse fitting. The results are compared with those of the LMedS method and the LTS method.

#### 5.1. Circle fitting

In Fig. 4, about 45% clustered outliers were added to the original circle data. Since LMedS and LTS only rely on the residuals of the data points, their results were affected by the standard variance of the inliers and percentages of the clustered outliers. Therefore, they failed to fit the circle under high percentage of clustered outliers (see Fig. 1). However, because the LTSD method considers the symmetry of the object, this enables LTSD find the right model (see Fig. 4): the true centre and radius of the circle are respectively  $(0.0, 0.0)$  and

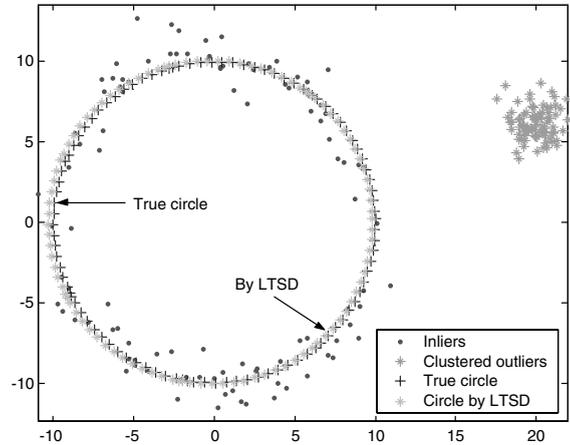


Fig. 4. Using the symmetry of the circle, the LTSD method found the approximately right results under 44.5% clustered outliers.

10.0; by the LTSD method, we obtained centre  $(-0.23, 0.01)$  and radius 10.06.

Another example showing the advantages of the proposed method is given in Fig. 5(a) (corresponding to Fig. 3(a)). From Fig. 5(a), we can see that when the outliers are clustered, the LMedS and LTS broke down under very low percentages of outliers, in this case, they both broke down under 38% outliers! In comparison to the LMedS and LTS methods, the proposed method gives the

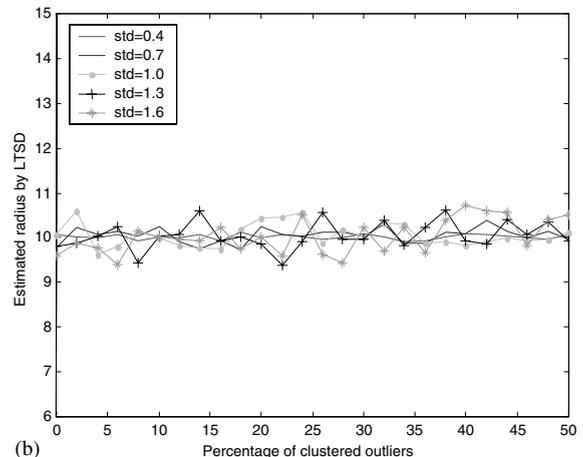
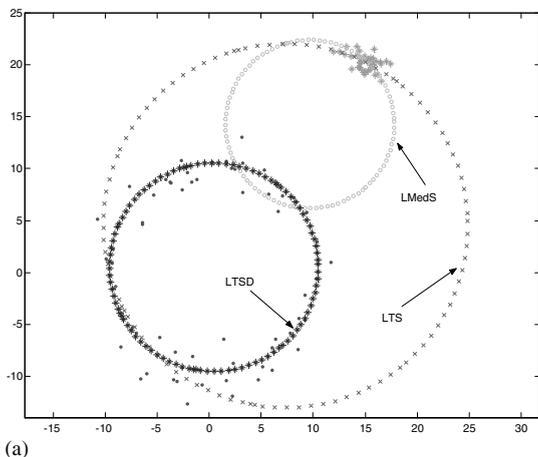


Fig. 5. (a) A comparative result of the LTSD, LMedS, and LTS with 38% clustered outliers and (b) the results of the LTSD method is affected less by the standard variance of inliers and percentages of clustered outliers.

most accurate results. The proposed method is affected less by the standard variance of the inliers and the percentages of the clustered outliers. Fig. 5(b) shows that the radius found by the LTSD method in circle fitting (true radius is 10.0) changed less under different standard variance of the inliers and percentages of clustered outliers. In comparison to Fig. 3(b) and (c), the fluctuation of the radius found by the LTSD method is smaller. Even when 50% clustered outliers exist in the data and the standard variance of inliers is 1.6, the results did not (yet) break down. However, both the LMedS and the LTS broke down.

### 5.2. Ellipse fitting

Ellipses are one of most common and important primitive models in computer vision and pattern recognition, and often occur in geometric shapes, man-made and natural scenes. Ellipse fitting is a very important task for many industrial applications because it can reduce the data and benefit the higher level processing (Fitzgibbon et al., 1999). Circles may be projected into ellipses under perspective projection. Thus ellipses are frequently used in computer vision for model matching (Sampson, 1982; Fitzgibbon et al., 1999; Robin, 1999, etc.). In this subsection, we apply the proposed robust method to ellipse fitting.

A general conic equation can be written as follows:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (5.1)$$

( $a, b, c, d, e, f$ ) are the parameters needed to find from the given data. When  $b^2 < 4ac$ , the equation above corresponds to ellipses.

The ellipse can also be represented by its more intuitive geometric parameters:

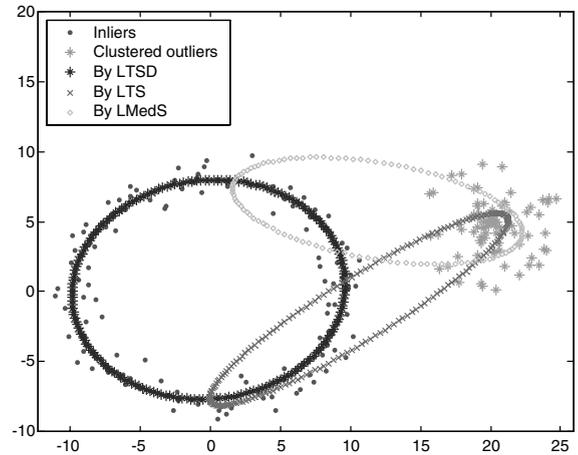


Fig. 6. Comparison of the results obtained by the LTSD method, LTS and LMedS in ellipse fitting under 40% clustered outliers.

$$\frac{(x \cos \theta + y \sin \theta - x_c \cos \theta - y_c \sin \theta)^2}{A^2} + \frac{(-x \sin \theta + y \cos \theta + x_c \sin \theta - y_c \cos \theta)^2}{B^2} = 1 \quad (5.2)$$

where ( $x_c, y_c$ ) is the center of the ellipse,  $\{A, B\}$  are the major and minor axes, and  $\theta$  is the orientation of the ellipse.

The relation between ( $a, b, c, d, e, f$ ) and ( $x_c, y_c, A, B, \theta$ ) is (Robin, 1999):

$$\begin{cases} x_c = \frac{be - 2cd}{4ac - b^2} \\ y_c = \frac{bd - 2ae}{4ac - b^2} \\ \{A, B\} = 2 \sqrt{\frac{-2f}{a + c \pm f \sqrt{b^2 + \left(\frac{a-c}{f}\right)^2}}} \\ \theta = \frac{1}{2} \tan^{-1} \frac{b}{a-c} \end{cases} \quad (5.3)$$

Table 1

Comparison of the estimated parameters by the LTSD, LTS, and LMedS methods in ellipse fitting under 40% clustered outliers

	$x_c$	$y_c$	Major axis	Minor axis	$\theta$ (°)
True value	0.0	0.0	10.0	8.0	0.0
The LTSD method	-0.125	-0.145	9.760	7.810	6.355
The LTS method	19.786	5.162	3.328	3.035	34.129
The LMedS method	9.560	5.208	11.757	3.679	-3.307

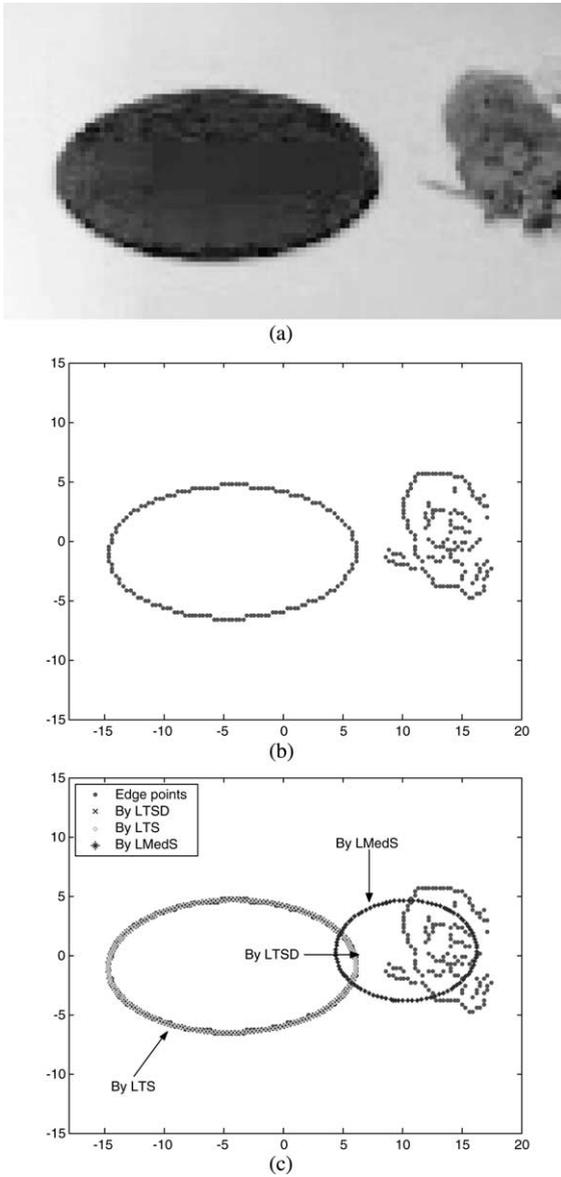


Fig. 7. Fitting a mouse pad: (a) a mouse pad with some flowers; (b) the edge image by using Canny operator and (c) the results obtained by the LTSD, LTS and LMedS methods.

It is convenient to find  $(a, b, c, d, e, f)$  first by the given data and then convert to  $(x_c, y_c, A, B, \theta)$ .

As illustrated in Fig. 6 and Table 1, 200 data were generated with 40% clustered outliers. The outliers were compacted within a region of radius 5 and center at  $(20.0, 5.0)$ . The ellipse had a standard variance 0.8, major axis 10.0, minor axis 8.0,

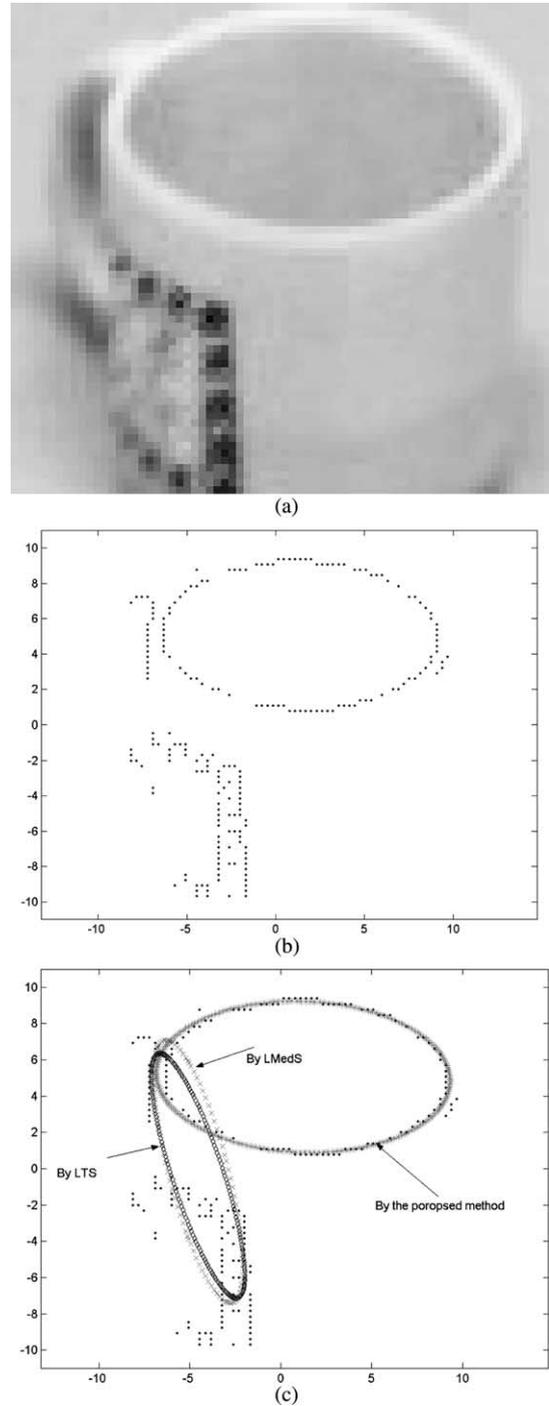


Fig. 8. Fitting the ellipse in a cup: (a) a real cup image; (b) the edge of the cup by applying Prewitt operator and (c) comparative results obtained by the LTSD, LTS and LMedS methods.

center (0.0,0.0), and orientation to horizon direction is  $0.0^\circ$ . The results of LTS and LMedS were seriously affected by the clustered outliers. However, the LTSD method worked well.

Next, we will apply the LTSD method to real images.

### 5.3. Experiments with real images

The first example is to fit an ellipse in an image of a mouse pad, shown in Fig. 7. The edge image was obtained by using Canny operator with threshold 0.07. In total, 310 data points were in the edge image (Fig. 7(b)). The clustered outliers, due to the flower, occupy 50% of the data. Three methods (the LTSD, LTS and LMedS) were applied to detect the mouse pad edge. As shown in Fig. 7(c), both LTSD and LTS correctly found the edge of the mouse pad. However, LMedS fails to detect the edge of the mouse pad. This is because under the condition that the standard variance of inliers is small, the statistical efficiency of LTS is better than LMedS.

Fig. 8 shows the use of the LTSD method to fit an ellipse to the rim of a cup. Fig. 8(a) gives a real cup image. After applying the Prewitt operator, the edge of the cup is detected is shown in Fig. 8(b). We can see there is a high percentage (about 45%) of clustered outliers existing in the edge image, external to the rim of the cup (the ellipse we shall try to fit), mainly due to the figure on the cup. However, the rim of the cup has a symmetric elliptical structure. Fig. 8(c) shows that the LTSD method correctly finds the ellipse in the opening of the cup, while both the LTS and the LMedS fail to correctly fit the ellipse.

### 5.4. Experiments for the data with uniform outliers

Finally, we investigated the characteristics of the LTSD under uniform outliers. We generated

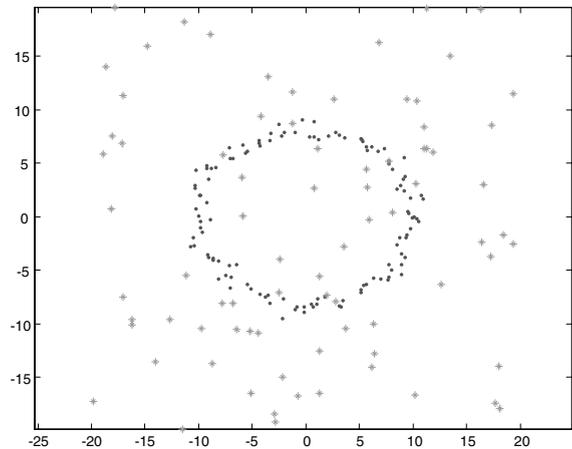


Fig. 9. An ellipse with 40% randomly distributed outliers.

200 data points with 40% uniform outliers (see Fig. 9). The ellipse had a standard variance 0.5, major axis 10.0, minor axis 8.0, center (0.0,0.0), and orientation to horizon direction  $\theta$  is  $0.0^\circ$ . The uniform outliers were randomly distributed in a rectangle with left upper corner  $(-20.0, 20.0)$  and right lower corner  $(20.0, -20.0)$ . We repeated the performance 100 times and the averaged results were shown in Table 2. We can see the LTSD method can also work well in uniform outliers.

## 6. Conclusion

The fragility of traditionally employed robust estimators: LMedS and LTS, in the presence of clustered outliers has been demonstrated in this paper (a similar story applies more widely—see Wang and Suter, 2003). These robust estimators can break down at surprisingly lower percentage of outliers when the outliers are clustered. Thus this paper provides an important cautionary note

Table 2

Comparison of the estimated parameters by the LTSD, LTS, and LMedS methods in ellipse fitting with 40% randomly distributed outliers

	$x_c$	$y_c$	Major axis	Minor axis	$\theta$ ( $^\circ$ )
True value	0.0	0.0	10.0	8.0	0.0
The LTSD method	-0.107	0.121	10.005	8.024	-1.119
The LTS method	-0.009	0.120	9.877	7.959	-2.117
The LMedS method	0.007	0.002	9.987	8.062	-0.981

to the computer vision community to carefully employ robust estimators when outliers are clustered. We also proposed a new method that incorporates SD into model fitting. The comparative study shows that this method can achieve better performance than the LMedS method and the LTS method especially when large percentages of clustered outliers exist in the data and the standard variance of inliers is large. The price paid for the improvement in fitting models is an increase of the computational complexity due to the complicated definition of symmetry distance. It takes about  $O(n^2)$  time to compute SD for each  $p$ -subset. The time complexity of the proposed method is also related to the times that how many  $p$ -subsets are needed (according to equation (2.7)). The proposed method can be applied to other symmetric shapes fitting and other fields.

Unfortunately, our method was especially designed for spatially symmetric data distributions. For inlier distributions that are not spatially symmetric (including structures that, though they may be symmetric, have large amounts of missing or occluded data so that the visible inliers are not symmetric), the LTSD is not a good choice. However, the LTSD does provide a feasible way to greatly improve achievements of conventional estimators—the LMedS and the LTS, especially, when the data contain inliers (with symmetry) with large variance and are contaminated by large percentage of clustered outliers.

Further work includes finding a more efficient and simple definition of SD by which the computational complexity can be reduced moderately. We will also extend our horizon to other features of the model studied so that the new method can be used more widely (data with different types of symmetry and reducing the limits required in that symmetry).

### Acknowledgements

This work is supported by the Australia Research Council (ARC), under the grant A10017082. The authors would like to thank reviewers for their valuable comments. The authors also thank Alireza Bab-Hadiashar for his helpful suggestions.

### References

- Attneave, F., 1995. Symmetry Information and Memory for Pattern. *Am. J. Psychol.* 68, 209–222.
- Bab-Hadiashar, A., Suter, D., 1998. Robust optical flow estimation. *Int. J. Comput. Vision* 29 (1), 59–77.
- Bigun, J., 1988. Recognition of local symmetries in gray value images by harmonic functions. In: *Proc. Int'l Conf. Pattern Recognition*, pp. 345–347.
- Danuser, G., Stricker, M., 1998. Parametric model fitting: From inlier characterization to outlier detection. *IEEE Trans. Pattern Anal.* 20 (2), 263–280.
- 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 (6), 381–395.
- Fitzgibbon, A., Pilu, M., Fisher, R.B., 1999. Direct least square fitting of ellipses. *IEEE Trans. Pattern Anal.* 21 (5), 476–480.
- Hampel, F.R., Rousseeuw, P.J., Ronchetti, E., Stahel, W.A., 1986. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, New York.
- Haralick, R.M., 1986. Computer vision theory: The lack thereof. *Comput. Vision Graph.* 36, 372–386.
- Hössjer, O., 1994. Rand-based estimates in the linear model with high breakdown point. *J. Am. Stat. Assoc.* 89, 149–158.
- Huber, P.J., 1973. Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Stat.* 1, 799–821.
- Huber, P.J., 1981. *Robust Statistics*. Wiley, New York.
- Kirby, M., Sirovich, L., 1990. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Anal.* 12 (1), 103–108.
- Kumar, R., Hanson, A.R., 1989. Robust estimation of camera location and orientation from noisy data having outliers. In: *Proc. Workshop on International 3D Scenes*, Austin, TX, pp. 52–60.
- Lee, K.-M., Meer, P., et al., 1998. Robust adaptive segmentation of range images. *IEEE Trans. Pattern Anal.* 20 (2), 200–205.
- Li, G., 1985. *Robust Regression, Exploring Data Tables, Trends and Shapes*. Wiley, New York.
- Marola, G., 1989. On the detection of the axes of symmetry of symmetric and almost symmetric planar Images. *IEEE Trans. Pattern Anal.* 11 (1), 104–108.
- Meer, P., Mintz, D., Rosenfeld, A., 1990. Least median of squares based robust analysis of image structure. In: *Proc. DARPA Image Understanding Workshop*, Pittsburgh, PA, pp. 231–254.
- Miller, W., 1972. *Symmetry Groups and Their Applications*. Academic Press, London.
- Ming, Y., Haralick, R.M., 2000. Optical flow from a least-trimmed squares based adaptive approach. In: *Proc. 15th International Conference on Pattern Recognition*, 3, pp. 1052–1055.
- Nalwa, V.S., 1989. Line-drawing Interpretation: Bilateral symmetry. *IEEE Trans. Pattern Anal.* 11 (10), 1117–1120.

- Reisfeld, D., Wolfson, H., Yeshurun, Y., 1992. Robust detection of facial features by generalized symmetry. In: Proc. Int'l Conf. Pattern Recognition, Champaign, Ill., pp. 117–120.
- Robin, P.L., 1999. Further five-point fit ellipse fitting. *Graph. Model Image Proc.* 61, 245–259.
- Roth, G., Levine, M.D., 1990. Segmentation of geometric signals using Robust fitting. In: Proc. 10th. Conf. on Pattern Recognition 1, pp. 826–831.
- Rousseeuw, 1984. Least median of squares regression. *J Am. Stat. Assoc.* 79, 871–880.
- Rousseeuw, P.J., Leroy, A., 1987. *Robust Regression and Outlier Detection*. John Wiley & Sons, New York.
- Rousseeuw, P.J., Aelst, S.V., 1999. Positive-Breakdown Robust Methods in Computer Vision. In: *Computing Science and Statistics*, vol. 31. Interface Foundation of North America, Inc. 451–460.
- Rousseeuw, P.J., Driessen, K.V., 1999. Computing LTS Regression for Large Data Sets, Technique Report, Univ. of Antwerp.
- Sampson, P.D., 1982. Fitting conic sections to very scattered data: An iterative refinement of the bookstein algorithm. *Comput. Graph. Image Proc.* 18, 97–108.
- Siegel, A.F., 1982. Robust regression using repeated medians. *Biometrika* 69, 242–244.
- Stewart, C.V., 1999. Robust parameter estimation in computer vision. *SIAM Rev.* 41 (3), 513–537.
- Su, M.C., Chou, C.H., 2001. A modified version of the K-means algorithm with a distance based on cluster symmetry. *IEEE Trans. Pattern Anal.* 23 (6), 674–680.
- Thomas, P.H., Simon, J.S., 1992. A cautionary note on the method of least median squares. *Am. Stat.* 46 (2), 79–83.
- Wang, H., Suter, D., 2003. MDPE: A Very Robust Estimator for Model Fitting and Range Image Segmentation, Technical Report (MECSE-2003-3), Monash University.
- Weyl, H., 1952. *Symmetry*. Princeton Univ. Press, Princeton, NJ.
- Yu, X., Bui, T.D., et al., 1994. Robust estimation for range image segmentation and reconstruction. *IEEE Trans. Pattern Anal.* 16 (5), 530–538.
- Zabrodsky, H., 1993. *Computational Aspects of Pattern Characterization—Continuous Symmetry*, PhD thesis, Hebrew Univ. Jerusalem, Israel.
- Zabrodsky, H., Peleg, S., Avnir, D., 1995. Symmetry as a continuous feature. *IEEE Trans. Pattern Anal.* 17 (12), 1154–1166.
- Zhang, Z., 1997. Parameter estimation techniques: A tutorial with application to conic fitting. *Image Vision Comput.* 15, 59–76.