Department of Electrical
and
Computer Systems Engineering

Technical Report
MECSE-8-2003

ASSC A New Robust Estimator for Data with Multiple Structures

Hanzi Wang and David Suter

MONASH
UNIVERSITY

ASSC: A New Robust Estimator for Data with Multiple Structures

Hanzi Wang and David Suter, Senior Member, IEEE
Department of Electrical and Computer Systems Engineering
Monash University, Clayton Vic. 3800, Australia.
{hanzi.wang ; d.suter}@eng.monash.edu.au

Abstract:

Estimating information from data with multiple structures has obtained more and more attention in computer vision community. When data include multiple structures, two major steps should be taken: i) robustly estimate the parameters of a model, and ii) differentiate inliers from outliers. In this paper, we propose two new robust techniques — robust Two-Step Scale estimator (TSSE) and robust Adaptive Scale Sample Consensus (ASSC) estimator. The first estimator (TSSE) applies nonparametric density estimation and density gradient estimation techniques, to robustly estimate the scale of inliers for heavily contaminated data. The second estimator (ASSC) is a complete robust fitting estimator. ASSC is based on both Random Sample Consensus (RANSAC) and TSSE. The ASSC estimator can tolerate more than 80% outliers. The main advantage of the ASSC estimator over RANSAC is that prior knowledge about the scale of inliers is not needed. The ASSC estimator can simultaneously estimate the parameters of a model and the scale of inliers belonging to that model. Comparative experiments show that the ASSC estimator has better robustness to heavily corrupted data with multiple structures than other robust methods: such as Least Median Squares (LMedS), Residual Consensus (RESC), and Adaptive Least Kth order Squares (ALKS).

1. Introduction

Multiple structures can occur in many computer vision problems such as range image segmentation, optical flow calculation, motion segmentation, etc. Great efforts have been made in the search for highly robust estimators that can deal with multiple-structural data in recent decades [2, 3, 16, 23, 26, 29]. A robust estimation technique is a method that can resist the influence of outliers ("bad data" and data associated with other structures are both called outliers in this paper). The least squares (LS) method can achieve optimal results under Gaussian distributed noise. However, this method is not robust because it is extremely sensitive to outliers (either 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 ([20], pp.9.). Because one single outlier is sufficient to force the LS estimator to produce an arbitrarily large value, the LS estimator has a breakdown point of 0%.

It is almost unavoidable that data are contaminated (due to faulty feature extraction, sensor noise, segmentation errors, etc) and it is also likely that the data will include multiple structures. Thus, there has recently been a general recognition that all computer vision algorithms should be robust [12]. When data include multiple structures, the structured data that are inliers to one model, can be outliers to another model. In contrast to uniformly distributed outliers and clustered outliers, these structured outliers are sometimes called *pseudo outliers* [24].

Although several robust estimators were developed during the past three decades, most of them (such as M-estimators [14], LMedS and LTS [18, 20], RM [21], etc) can only tolerate up to 50% gross errors. These traditional robust estimators assume that inliers occupy the absolute majority of the whole data. However, this assumption is far from being satisfied for the real tasks faced in computer vision [26]. In computer vision tasks, it frequently happens that gross noise and pseudo outliers occupy the absolute majority of the data. For example, when data include two structures (lines) and randomly distributed noise (see figure 5 (b)-(d)), no single structure occupies an absolute majority of the whole data set. In this case, traditional methods will break down because their basic assumption (i.e., inliers occupy the absolute majority) is violated. For this kind of case, we need to find a more robust estimator that can tolerate more than 50% outliers.

Some robust techniques: Hough Transform [13, 15] and RANSAC [10] were developed in the vision community. They need user to provide certain crucial parameters: a priori bin size (for Hough Transform) or priori error tolerance (for RANSAC). The bin size and the error tolerance are related to the scale of inliers. Thus, given the correct scale, both techniques can achieve good results for data with more than 50% outliers. However, a priori knowledge about the scale of the inlier is not often available. If the scale is wrongly provided, these two methods will fail. This problem greatly limits their application in many computer vision tasks.

RESC is one successful example of highly robust methods [29] and it can tolerate more than 80% outliers. The RESC method uses a compressed histogram method to infer residual consensus. Instead of using the size of the residuals as its criteria, the RESC method uses the “histogram power” as its criteria. The RESC method finds the parameters by choosing the p-subset

corresponding to the maximum histogram power. Unfortunately, the method needs the user to tune many parameters in compressing histogram. We also found the RESC method, after finding the parameters of a model, overestimates the scale of inliers (see section 2.4).

MINPRAN is another kind of estimator that claims to tolerate more than 50% outliers [23]. It can find the correct model in the data involving more than 50% outliers without a priori knowledge about error bounds and it claims that it does not " hallucinate" fits when there are no real structures in the data. However, MINPRAN assumes that the outliers are randomly distributed within a certain range. This makes MINPRAN less effective in extracting multiple structures.

The techniques MUSE [17] and ALKS [16] incorporate robust scale estimation, and both methods can arguably tolerate outliers in excess of 50%. MUSE and ALKS can perform better than LMedS and M-estimators at small-scale discontinuities. However, MUSE needs a lookup table for the scale estimator correction; ALKS is limited in its ability to handle extreme outliers. Another problem we found in ALKS is its lack of stability under a small percentage of outliers [26].

Bab-Hadiashar and Suter [1] have used least K-th order (rather than median) methods and a heuristic way of estimating scale to perform range segmentation. However, though their method can handle large percentages of outliers and pseudo-outliers, it does not seem to be successful in tolerating extreme cases.

Having a correct scale of inliers is crucial to the robust behaviour of an estimator. The success of some robust estimators is based on having correct initial scale estimate, or the correct setting of a particular parameter that is related to scale (e.g., RANSAC, Hough Transform, M-estimators etc.). However, their performance crucially depends on that user-provided scale-related knowledge. Robust scale estimation is often attempted during a postprocessing stage of robust estimators (such as LMedS, LTS, etc.). Yet, although there are a lot of papers that propose robust estimators with high breakdown point for model fitting, robust scale estimation is relatively neglected. We will, in this paper, investigate the behaviour of several state-of-the-art robust scale estimators for data with multiple structures, and propose a novel robust scale estimator (TSSE).

Furthermore, based on TSSE and RANSAC, we will propose a new robust estimator — called Adaptive Scale Sample Consensus (ASSC) estimator. The ASSC estimator is very robust for data

involving high percentages of outlier and multiple structures. It can tolerate more than 80% outliers.

The main contributions of this paper can be summarized as follows:

- We investigate robust scale estimation and propose a novel and effective robust scale estimator: Two-Step Scale Estimator (TSSE), based on nonparametric density estimation and density gradient estimation techniques.
- Based on RANSAC and TSSE, we propose a highly robust estimator: Adaptive Scale Sample Consensus (ASSC) estimator. The ASSC estimator is an important improvement over RANSAC because no priori knowledge concerning the scale of inliers is necessary (the scale estimation is data driven). ASSC can tolerate more than 80% outliers and multiple structures.
- Comparative experiments show that both the TSSE and ASSC estimators are highly robust to heavily corrupted data with multiple structures and they outperform the competing methods.

This paper is organized as follows: in section 2, we review previous robust scale techniques. In section 3, the density gradient estimation and mean shift/mean shift valley method are introduced, and the robust scale estimator: TSSE is proposed. TSSE is experimentally compared with five other robust scale estimators, using data with multiple structures, in section 4. The robust ASSC estimator is proposed in section 5 and comparative experimental results for 2D and 3D examples are contained in section 6. We conclude in section 7.

2. Robust scale estimators

The emphasis in most past work presenting robust estimators was on the high breakdown point [16, 17, 20, 23, 29], i.e. the estimator that can correctly find the parameters of a model from the data which are heavily contaminated. Whether or not the inliers can be successfully differentiated from the outliers depends on two factors: (1) whether the parameters of a model are correctly found; and (2) whether the scale of inliers is correctly estimated. Step (2), scale estimation plays an important role in the overall robust behavior of these methods. Some robust estimators, such as M-estimators, RANSAC, Hough Transform, etc., put the onus on the "user" - they simply require some user-set parameters that are linked to the scale of inliers. Others, such as LMedS, RESC, MDPE [26],, etc., use an auxiliary estimate of scale (after finding the parameters of a model) during a post-processing stage, which aims to differentiate inliers from outliers.

In this section, we will review several state-of-the-art robust scale estimators.

2.1 The Median and Median absolute deviation (MAD) scale estimator

Among many robust estimators, the sample median is one of the most famous estimators. The sample median is bounded when the data include more than 50% inliers. A robust median scale estimator is then given by [20]:

$$M = 1.4826(1+\frac{5}{n-p})\sqrt{med_i\, x_i^2} \tag{2.1}$$

where x_i is the residual of i'th sample, n is the number of sample points and p is the dimensions of parameter space (e.g., 2 for a line, 3 for a circle).

The inliers are usually taken to the data points that satisfy the following condition:

$$\left|\mathrm{xj/M}\right| < T \tag{2.2}$$

where T is a threshold. T is usually set to 2.5.

A variant, MAD, is also used to estimate the scale of inliers [19]:

$$\mathrm{MAD}=1.4826\mathrm{med}_i\{|x_i-\mathrm{med}_j x_j|\} \tag{2.3}$$

The MAD estimator is very robust to outliers and has a 50% breakdown point. The outliers can be recognized by computing:

$$\frac{\left|x_i - med_j x_j\right|}{MAD_n} \tag{2.4}$$

When equation (2.4) for a point x_i drawn from a sample $\{x_j\}$ exceeds a threshold, say 2.5, an outlier is recognized.

The median and MAD are often used to yield initial scale values (before estimating the parameters of a model) for many robust estimators. These two methods can also serve as auxiliary scale estimators (after finding the parameters of a model) for other robust estimators.

Because the median and MAD have 50% breakdown points, they will break down when the data include more than 50% outliers. Both methods are biased for multiple-mode cases even when the data contains less than 50% outliers (see section 4).

2.2 Adaptive Least K-th Squares (ALKS) Estimator

The authors of ALKS [16] consider robust k scale estimation and they search for a model by randomly choosing p-subsets and minimizing the k-th order statistics of the squared residuals. The robust k scale estimate, assuming inliers have a Gaussian distribution, is given by:

$$\hat{s}_k = \frac{\hat{d}_k}{\Phi^{-1}[(1+k/n)/2]} \tag{2.5}$$

where $\hat{d}_k$ is the half-width of the shortest window including at least k residuals; $\Phi^{-1}[\cdot]$ is the argument of the normal cumulative density function.

The optimal value of the k is that corresponds to the minimum of the variance of the normalized error ε_k^2:

$$\varepsilon_k^2 = \frac{1}{k-p}\sum_{i=1}^{k}\left(\frac{r_{i,k}}{\hat{s}_k}\right)^2 = \frac{\hat{\sigma}_k^2}{\hat{s}_k^2} \tag{2.6}$$

The authors assume that when k is increased so that the first outlier is included, the increase of $\hat{s}_k$ is much less than that of $\hat{\sigma}_k$.

ALKS is limited in its ability to handle extreme outliers. Another problem we found [26] in ALKS is its lack of stability under a small percentage of outliers.

2.3 Modified Selective Statistical Estimator (MSSE)

Bab-Hadiashar and Suter [1] use the least k-th order (rather than median) methods and a heuristic way of estimating scale to perform range segmentation. After finding a fit, they try to recognize the first outlier, corresponding to where the k-th residual "jumps", by looking for a jump in the unbiased scale estimate formed by using the first k-th residuals in an ascending order:

$$\hat{\sigma}_k^2 = \frac{\sum_{i=1}^{k} r_i^2}{k-p} \tag{2.7}$$

where p is the dimension of the model.

They assume that when k is increased, the value of the k-th residual will jump when it comes from a different distribution, leading to a corresponding jump in σ_k. Thus, emphasis is shifted from using a good scale estimate to define the outliers, to finding the point of breakdown in the unbiased scale estimate (thereby signaling the inclusion of an outlier) as defined by the first k that satisfies the following inequality:

$$\frac{\sigma_{k+1}^2}{\sigma_k^2} > 1 + \frac{T^2 - 1}{k - p + 1} \tag{2.8}$$

Because this method does not rely on the k-th order statistics *for the scale estimate* (it uses only the first k data points that has been classified as inliers), it is unbiased even when the data include multiple-structures.

However, though their method can handle large percentages of outliers and pseudo-outliers, it does not seem as successful in tolerating extreme cases.

2.4 Residual Consensus (RESC) Method

RESC is another successful example of a recent robust method [29]. After finding a fit, it estimates the scale of the fit by directly calculating:

$$\sigma = \alpha\left(\frac{1}{\sum_{i=1}^{v} h_i^c - 1} \sum_{i=1}^{v} (i h_i^c \delta - \bar{h}^c)^2\right)^{1/2} \tag{2.9}$$

where $\bar{h}^c$ is the mean of all residuals included in the compressed histogram; α is a correct factor for the approximation introduced by rounding residuals in a bin of histogram to $i\delta$ (δ is the bin size of the compressed histogram); v is the number of bins of the compressed histogram.

However, we find that the estimated scale is overestimated for the reason that, instead of summing up squares of the differences between all individual residuals and the mean residual in the compressed histogram, equation (2.9) sums up the squares of the differences between residuals in each bin of compressed histogram and the mean residual in the compressed histogram.

To reduce this problem, we propose an alternative form:

$$\sigma = \alpha\left(\frac{1}{\sum_{i=1}^{v} h_i^c - 1} \sum_{i=1}^{n_c} (r_i - \bar{h}^c)^2\right)^{1/2} \tag{2.10}$$

where n_c is the number of data points in the compressed histogram.

3. A robust scale estimator: TSSE

In this section, we will introduce some density estimation and density gradient estimation techniques. Then, we propose a highly robust scale estimator (TSSE), which is very robust to multiple-structural data.

3.1 Density Gradient Estimation and Mean Shift Method

There are several nonparametric methods available for probability density estimation: the histogram method, the naive method, the nearest neighbor method, and kernel estimation [22] The kernel estimation method is one of the most popular techniques. Given a set of n data points $\{X_i\}_{i=1,\ldots,n}$ in a d-dimensional Euclidian space R^d, the multivariate kernel density estimator with kernel K and window radius (band-width) h is defined as follows ([22], p.76)

$$\hat{f}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K(\frac{x - X_i}{h}) \tag{3.1}$$

The kernel function K(x) should satisfy some conditions ([25], p.95).

There are several different kernels. The Epanechnikov kernel ([22], p.76) is one optimum kernel which yields minimum mean integrated square error (MISE):

$$K_e(\mathrm{X}) = \begin{cases} \frac{1}{2} c_d^{-1}(d+2)(1-\mathrm{X}^T\mathrm{X}) & if\ \mathrm{X}^T\mathrm{X} < 1 \\ 0 & otherwise \end{cases} \tag{3.2}$$

where c_d is the volume of the unit d-dimensional sphere, e.g., $c_1=2$, $c_2=\pi$, $c_3=4\pi/3$.

The estimate of the density gradient can be defined as the gradient of the kernel density estimate (3.1)

$$\hat{\nabla} f(x) \equiv \nabla \hat{f}(x) = \frac{1}{nh_d}\sum_{i=1}^{n} \nabla K(\frac{x - X_i}{h}) \tag{3.3}$$

According to (3.3), the density gradient estimate of the Epanechnikov kernel can be written as

$$\hat{\nabla} f(x) = \frac{n_x}{n(h^d c_d)} \frac{d+2}{h^2} \left(\frac{1}{n_x} \sum_{X_i \in S_h(x)} [X_i - x] \right) \tag{3.4}$$

where the region $S_h(x)$ is a hypersphere of the radius h, having the volume $h^d c_d$, centered at x, and containing n_x data points.

The mean shift vector $M_h(x)$ is defined as

$$M_h(x) \equiv \frac{1}{n_x} \sum_{X_i \in S_h(x)} [X_i - x] = \frac{1}{n_x} \sum_{X_i \in S_h(x)} X_i - x \tag{3.5}$$

Equation (3.4) can be rewritten as

$$M_h(x) \equiv \frac{h^2}{d+2} \frac{\hat{\nabla} f(x)}{\hat{f}(x)} \tag{3.6}$$

Equation (3.6) firstly appeared in [11]. Equation (3.5) shows that the mean shift vector is the difference between the local mean and the center of the window. Equation (3.6) shows the mean shift vector is an estimate of the normalized density gradient. The mean shift is an unsupervised nonparametric estimator of density gradient. One characteristic of the mean shift vector is that it always points towards the direction of the maximum increase in the density.

The Mean Shift algorithm can be described as follows:

1. Choose the radius of the search window and initialize the location of the window.
2. Compute the mean shift vector $M_h(x)$.
3. Translate the search window by $M_h(x)$.
4. Step 2 and step 3 are repeated until convergence.

The converged centers (or windows) correspond to modes (or centers of the regions of high concentration) of data represented as arbitrary-dimensional vectors. Since its introduction by Fukunaga and Hostetler [11], the mean shift method has been extensively exploited and applied in low level computer vision tasks [4-7] for its ease and efficiency.

Almost all published methods, which employ the mean shift method, use its hill climbing property to find the peaks of some functions of interest (residual pdf's in our case). However, sometimes it is very important to find the valleys. Two recently published papers investigate this in the context of mean shift [8, 27]. Although the employed kernel in these two papers is different (the Gaussian kernel in [8]; the Epanechnikov kernel in [27]) and the name of the local minimum density point is different (called saddle point in [8]; and valley point in [27]), both methods seek for the local minimum density. In next subsection, we will summarize a simple mean shift valley algorithm [27].

3.2 Mean Shift Valley Algorithm

One characteristic of the mean shift vector is that it always points towards the direction of the maximum increase in the density. Thus the opposite direction of the mean shift vector will always points toward to a local minimum density.

In order to find valley in density space, we define the mean shift valley vector:

$$\mathrm{MV_h}(x) = \text{-}\mathrm{M_h}(x) = x - \frac{1}{n_x} \sum_{x_i \in S_h(x)} x_i \tag{3.7}$$

Replace $\mathrm{M_h}(x)$ in (3.6) by $\mathrm{MV_h}(x)$, we can obtain:

$$\mathrm{MV_h}(x) \equiv -\frac{h^2}{d+2} \frac{\hat{\nabla} f(x)}{\hat{f}(x)} \tag{3.8}$$

$\mathrm{MV_h}(x)$ always points towards the direction of the maximum decrease in the density.

In practice, the step-size given by the above analysis may lead to oscillation. We have observed this case particularly when finding valleys (although the potential exists when seeking peaks as well). Thus we derive a recipe for avoiding the oscillations in valley seeking.

Let $\{y_k\}_{k=1,2\ldots}$ be the sequence of successive locations of the mean shift valley procedure, then we have, for each k=1,2…,

$$y_{k+1} = y_k + p \cdot MV_h(y_k) \tag{3.9}$$

p is a correction factor, and $0 < p \le 1$. If the shift step at y_k is too large, it causes y_{k+1} to jump over the local valley and thus oscillate over the valley. This problem can be avoided when we adjust the correction factor p so that $\mathrm{MV_h}(y_k)^{\mathrm{T}}\, \mathrm{MV_h}(y_{k+1}) > 0$.

The mean shift valley algorithm can be described as:

1. Choose the radius of the search window, set p =1, and initialize the location of the window
2. Compute the shift step vector $\mathrm{MV_h}(y_k)$.
3. Compute y_{k+1} by equation (3.9) and $\mathrm{MV_h}(y_{k+1})$.
4. If $\mathrm{MV_h}(y_k)^{\mathrm{T}}\, \mathrm{MV_h}(y_{k+1}) > 0$, go to step 5; Otherwise, we let $p=p/2$. Repeat step 3 and 4 until $\mathrm{MV_h}(y_k)^{\mathrm{T}}\, \mathrm{MV_h}(y_{k+1}) > 0$;
5. Translate the search window by $p \cdot MV_h(y_k)$.
6. Repeat step 2 to step 5 until convergence.

There is one exceptional case: when there are no local valleys (e.g., uni-modal), the mean shift valley method is divergent. This can be easily avoided by terminating when no samples fall within the window.

3.3 Bandwidth Choice

One crucial issue in the non-parametric density estimation, in the mean shift, and in mean shift valley methods, is how to choose bandwidth h. There are a lot of papers discussing the choice of bandwidth [7, 9, 25]. Because we apply the mean shift method and estimate the density in one-dimensional residual space, a simple over-smoothed bandwidth selector can be employed [25].

$$\hat{h} = \left[\frac{243R(K)}{35u_2(K)^2 n} \right]^{1/5} S \tag{3.10}$$

where $R(K) = \int_{-1}^{1} K(\zeta)^2 d\zeta$ and $u_2(K) = \int_{-1}^{1} \zeta^2 K(\zeta) d\zeta$. S is the sample standard deviation.

The median, MAD or robust k scale estimator can be used to yield an initial scale estimate. $\hat{h}$ will provide an upper bound on the AMISE (asymptotic mean integrate error) optimal bandwidth $\hat{h}_{AMISE}$. The median, MAD, and robust scale estimator may be biased for data with multi-model case. This is because these estimators are proposed assuming the whole data have a Gaussian distribution. Because the bandwidth in equation (3.10) is proportional to the estimated scale, the bandwidth can be set as $c\hat{h}$, where c is a correct factor (0<c<1) and is used to avoid over-smoothing ([25], p.62).

To illustrate the mean shift and mean shift valley method, three normal modes (mode 1 includes 600 data points with mean 0.0, mode 2 includes 500 data points with mean 4.0, and mode 3 includes 600 data points with mean 8.0) with total 1700 data points were generated (see figure 1). To find local peaks, we set two initial points: P0 (-2.0) and P1(5.0). After applying the mean shift method, we obtained the two local peaks: P0'(0.01) and P1'(4.03). Similar, we applied the mean shift valley method to find the local valleys. We selected two initial points: V0 (0.5) and V1 (7.8). The mean shift valley method automatically found the local minimum densities. Precisely, V0' was located at 2.13, and V1' was at 6.00.

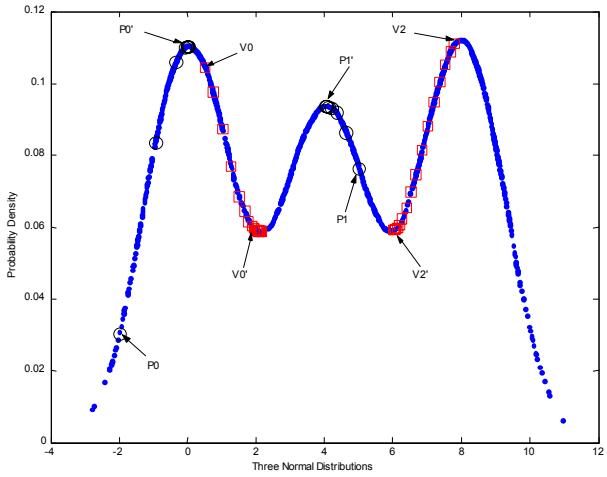

Fig. 1. An example of applying the mean shift method to find local peaks and applying the mean shift valley method to find local valleys.

3.4 Two-step scale estimator (TSSE)

We base our method of scale estimation on the assumption that the inliers occupy the relative majority, and are Gaussian or Gaussian-like distributed, but the whole data can include multi-structural distributions. Thus, we propose a robust two-step method to estimate the scale of the inliers.

(1) Because the inliers have Gaussian or Gaussian-like distribution, we use mean shift, with initial center zero, to find the local peak, and then we use the mean shift valley to find the valley next to the peak. All these are performed in ascending ordered absolute residual space. Thus, modes other than the inliers will be disregarded as they lie outside the obtained valley.

(2) We estimate the scale of the fit by the median scale estimator on the points within the obtained window centered at the local peak.

TSSE is very robust to outliers and can resist heavily contaminated data with multiple structures. In next section, we will compare the achievements of our method and other five methods. The experiments will show the advantages of the proposed method over other methods.

4. Experimental comparisons on scale estimation

In this section, we will investigate the behavior of several state-of-the-art robust scale estimators that are widely used in computer vision community and show the weakness of these scale

estimation techniques. We assume we know the parameters of the model. In the following experiments, we compare the proposed method — TSSE, with other five robust scale estimators: the median, MAD, ALKS, MSSE, and the revised RESC (according to equation 2.10). Comparative experiments show the proposed method achieve better results than the other five robust scale estimators.

The signals were generated as follows: The i'th structure has γ_i data points, corrupted by Gaussian noise with zero mean and standard variance σ_i. α data points were randomly distributed in the range of (0, 100).

4.1 Normal distribution

First, we generate a simple line signal: One line: x:(0-55), y=30, γ_1=10000, σ_1=3; α=0, i.e., 100% inliers; After we applied the six robust scale estimators to the signal, we obtained the median (3.0258); MAD (3.0237); ALKS (2.0061); MSSE (2.8036); the revised RESC (2.8696); and TSSE (3.0258). Among these six comparative methods, the median, MAD, and TSSE gave the most accurate results. ALKS gave the worst result. This is because the robust estimate $\hat{s}_k$ is an underestimate of σ for all values of k (17, p.202) and because the criterion (2.6) estimates the optimal k wrongly. ALKS used only about 15% data as inliers. MSSE used 98% data points as inliers, which is reasonably good.

4.2 Two-mode distribution

In this subsection, we use relatively complicated data. We generated a step signal so that the data include two structures, i.e. two lines.

A step signal: line1: x:(0-55), y=40, γ_1=3000, σ_1=3; line2: x:(55-100), y=70, γ_2=2000, σ_2=3; α=0.

The results that we obtained are as follows: the median (6.3541); MAD (8.8231); ALKS (3.2129); MSSE (2.8679); the revised RESC (2.9295); and TSSE (3.0791). Among these six comparative methods, the median and MAD gave the worst results. This is because the median and MAD scale estimators assume the residuals of the whole data are at Gaussian distribution, which is violated in the signal (containing two modes). The other four robust scale estimators yield good results.

4.3 Two-mode distribution with random outliers

Next, we again use the above one-step signal. However, we increased the number of outliers so that the data include 80% of outliers, i.e., γ_1=1000; γ_2=750; α=3250.

After applying the six methods, the estimated scale of the signal that we obtained are: the median (34.0962); MAD (29.7909); ALKS (7.2586); MSSE (27.4253); the revised RESC (24.4297); and TSSE (4.1427). From the obtained results, we can see that only the proposed method gave a reasonably good result, while all other five methods failed to estimate the scale of the inliers when the data involve a high percentage of outliers.

4.4. Breakdown plot

4.4.1 A roof signal

We generate a roof signal containing 500 data points in total.

A roof: x:(0-55), y=x+30, γ_1, σ=2; x:(55-100), y=140-x, γ_2=50; σ=2.

At the beginning, we assign 450 data point to γ_1 and the number of the uniform outliers α =0; Thus, the data include 10% outliers. Then, we decrease γ_1, and at the same time, we increase α so that the total number of data points is 500. Finally, γ_1=75, and α=375, i.e. the data include 85% outliers. The results are repeated 20 times.

Figure 2 shows that TSSE yielded the best results among the six comparative methods. The revised RESC method begins to break down when the outliers occupy around 60%. MSSE gave reasonable results when the percentage of outliers is less than 75%, but it broke down when the data include more outliers. Although the breakdown points of the median and the MAD scale estimators are as high as 50%, their results deviated from the true scale even when outliers are less than 50% of the data. They are biased more and more from the true scale with the increase in the percentage of outliers. ALKS yielded less accurate results than TSSE, and less accurate results than the revised RESC and MMSE when outliers are less 60%.

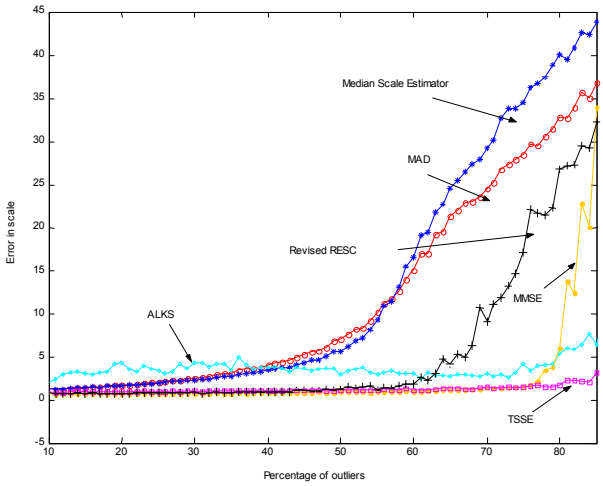

Fig. 2. Breakdown plot of six methods in estimating the scale of a roof signal.

4.4.2 A step signal

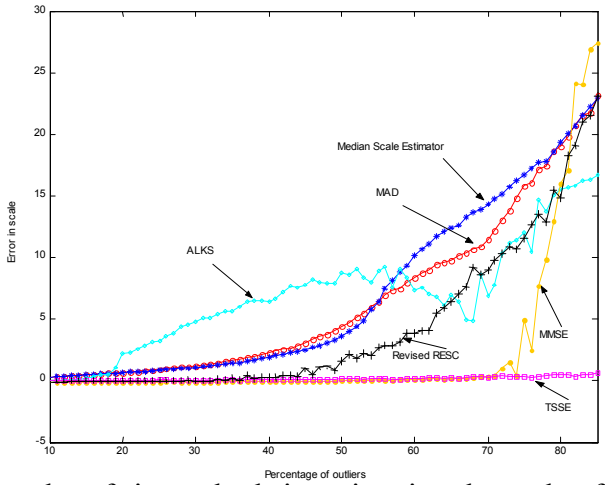

Fig. 3. Breakdown plot of six methods in estimating the scale of a step signal.

We generated another signal: one-step signal that contains 1000 data points in total.

One-step signal: x:(0-55), y=30, γ_1, σ=2; x:(55-100), y=40, γ_2=100; σ=2.

At the beginning, we assign γ_1 900 data points and the number of the uniform outliers α =0; Thus, the data include 10% outliers. Then, we decrease γ_1, and at the same time, we increase α so that the number of the whole data points is 1000. Finally, γ_1=150, and α=750, i.e. the data include 85% outliers.

From figure 3, we can see TSSE gave the most accurate estimation of the scale of the signal. In contrast, the revised RESC begins to break down when the number of outliers is about 50% of the data. MSSE gave reasonable results when the percentage of outliers is less than 70%. However, it broke down when the data include more outliers. The median and the MAD scale estimators are more and more biased with the increase in the percentage of outliers for the two-structured signal. ALKS yielded less satisfactory results.

Compared with figure 2, we can see that the revised RESC, MSSE, and ALKS yielded less accurate results for small scale step signal than roof signal, but the results of the proposed TSSE are similar accurate for both types of signals. Even when the data include 85% outliers, the recovered scales of inliers by TSSE for the one-step signal are 2.95, which is reasonably good.

4.4.3 Breakdown plot for robust k scale estimator

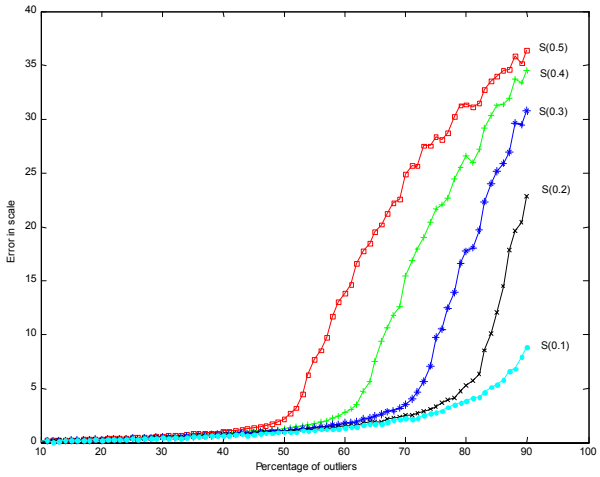

Fig.4 Breakdown plot of different robust scale estimator

If the data have a Gaussian like distribution, the median scale estimator (2.1) is only one possible robust k scale estimator (2.5). We investigated the achievements of the robust k scale estimator after the correct parameters of a model have been found. Let:

$$S(q) = \frac{\hat{d}_q}{\Phi^{-1}[(1+q)/2]} \tag{4.1}$$

where q is set from 0 to 1. Thus S(0.5) is the median scale estimator.

We generated a one-step signal containing 500 data points in total.

One-step signal: x:(0-55), y=30, γ_1, σ=1; x:(55-100), y=40, γ_2=50; σ=1.

At the beginning, γ_1 = 450 and α =0; Then, we decrease γ_1, and at the same time, we increase α until γ_1=50, and α=400, i.e. the data include 90% outliers.

Consider the dual issues of parameter estimation and sale estimation. As figure 4 shows, after finding the robust estimate of the parameters of a model, the accuracy of S(q) is increased with the decrease of q. When the outliers are less than 50% of the whole data, the difference for different values of q is small. However, when the data include more than 50% outliers, the difference for various values of q is large. This provides a useful cue for robust estimators, which use the median scale method to recovery the scale of inliers.

From the experiments in this section, we can see the proposed TSSE is a very robust scale estimator, and it achieves better results than the other five comparative methods when the parameter of a model is known. In practice, however, the parameters of a model are to be estimated. When the parameters of a model are unknown, we need a robust estimator to estimate the parameters. In the next section, we will propose a new robust estimator—Adaptive Scale Sample Consensus (ASSC) estimator, which can estimate the parameters and the scale simultaneously.

5. Robust Adaptive Scale Sample Consensus Estimator

Fischler and Bolles [10] provided a generate-and-test paradigm: RANdom Sample Consensus (RANSAC). They used the minimum number of data points, a p-subset (p is the dimension of parameter space, e.g., p=2 for line fitting, p=3 for plane fitting), necessary to estimate the parameters θ of the model. Like LMedS, RANSAC use a random sampling technique, randomly sample p-subsets from the whole data until at least one p-subset is clean. A p-tuple is "clean" if it consists of p good observations without contamination by outliers. One performs m times the random selections of p-tuples, where m is chosen so that the probability P that at least one of the m p-tuples is "clean" is almost 1. Let ε be the fraction of outliers contained in the whole set of points. The probability P can be expressed as follows ([20], pp.198):

$$P=1-(1-(1-\varepsilon)^{p})^{m} \tag{5.1}$$

Thus one can determine m for given values of ε, p and P by:

$$m=\frac{\log(1-P)}{\log[1-(1-\varepsilon)^{p}]} \tag{5.2}$$

The criterion of the RANSAC method is to maximize the number of data points within the user-set error bound. Clearly, this bound is related to the scale of the inliers (S). Mathematically, the RANSAC estimate can be written as:

$$\hat{\theta}=\arg\max_{\hat{\theta}} n_{\hat{\theta}} \tag{5.3}$$

where $n_{\hat{\theta}}$ is the number of points whose absolute residual in the candidate parameter space is within the error bound (i.e., $|r| \le 2.5S$)); $\hat{\theta}$ is the estimated parameters from one of the randomly chosen p-subsets.

The error bound in RANSAC is crucial to the achievements of RANSAC. Provided with a correct error bound of inliers, the RANSAC method can find a model even when data contain a large percentage of gross errors. However, when the error bound is wrongly given, RANSAC will totally break down even when outliers occupy small percentages of the whole data [26]. Thus the major problem with RANSAC is that the technique needs priori knowledge of the error bound of inliers, which is not available in most practical vision tasks.

5.1 The algorithm of Adaptive Scale Sample Consensus Estimator

We assume that when a model is correctly found, two criterions should be satisfied:

1. The number of data points (n_θ) near or on the model should be as large as possible;
2. The residuals of the inliers should be as small as possible. Correspondingly, the scale (S_θ) should be as small as possible.

RANSAC only considers the first criterion in its objective function. Our objective function considers both criteria. Thus, we define our objective function as n_θ/S_θ. The ASSC estimator can be written as:

$$\hat{\theta} = \arg\max_{\hat{\theta}}(n_{\hat{\theta}} / S_{\hat{\theta}}) \qquad (5.4)$$

The improvements of the proposed ASSC estimator over RANSAC are:

- In the objective function of the proposed method, both the number of data points within an error bound and the corresponding scale are considered. While in the RANSAC method, only the first criterion is considered.
- No priori knowledge about the scale of inliers is necessary in the proposed method. The robust ASSC estimator yields the estimated parameters of a model and the corresponding scale simultaneously.

The ASSC estimator is an extended version of RANSAC. When the estimate of the scale is fixed, equation 5.4 is another form of RANSAC with the score n_θ scaled by 1/S (i.e, a fixed constant for all p-subsets). Because RANSAC estimates the parameters of a model by finding the relative maximum score, the scaling in the score n_θ does not affect the results. ASSC is more reasonable because the scale is different for each candidate fit determined by each p-subset.

The procedure of the ASSC estimator algorithm is as follows:

(1) Randomly choose one p-subset from the data points, estimate the model parameters using the p-subset, and calculate the ordered absolute residuals of all data points.

(2) Choose bandwidth by equation 3.10. Robust k scale estimator (k=0.2) is used to yield a coarse initial scale.

(3) Apply TSSE to to the absolute sorted residuals to estimate the scale of inliers. At the same time, the probability density at the local peak $\hat{f}(peak)$ and local valley $\hat{f}(valley)$ are obtained by equation (3.1).

(4) Validate the valley. Let $\hat{f}(valley)/\hat{f}(peak) = \lambda$ (where $1>\lambda \geq 0$). Because the inliers are assumed having a Gaussian like distribution, the valley is invalid when λ is too large (say, 0.8). If the valley is valid, go to step (5); otherwise go to step (1).

(5) Calculate the score, i.e., the objective function of the ASSC estimator.

(6) Repeat step (1) to step (5) m times (m is set by equation 5.2). Finally, output the parameters and the scale S_1 with the highest score.

Because the robust k scale estimator is biased for data with multiple structures, the final scale of inliers S_2 can be refined when the scale S_1 obtained by TSSE is used. The results are from one p-subset, corresponding to the highest score. In order to improve the statistical efficiency, a weighted least square procedure ([20], p.202) can be carried out after finding the initial fit.

Instead of estimating the fit involving the absolute majority in the data set, the ASSC estimator finds a fit having a relative majority of the data points. This makes it possible, in practice, for the ASSC estimator to obtain a high robustness that can tolerate more than 50% outliers. The experiments in the next section show that the ASSC estimator is a very robust estimator for data with multiple structures and high percentages of outliers.

6. Experiments for data with multiple structures

In this section, both 2D and 3D examples are given to illustrate the robustness of the proposed method to data including a high percentage of outliers and multiple structures. The results of the proposed method are also compared with those of three other popular methods: LMedS, RESC, and ALKS.

6.1 2D examples

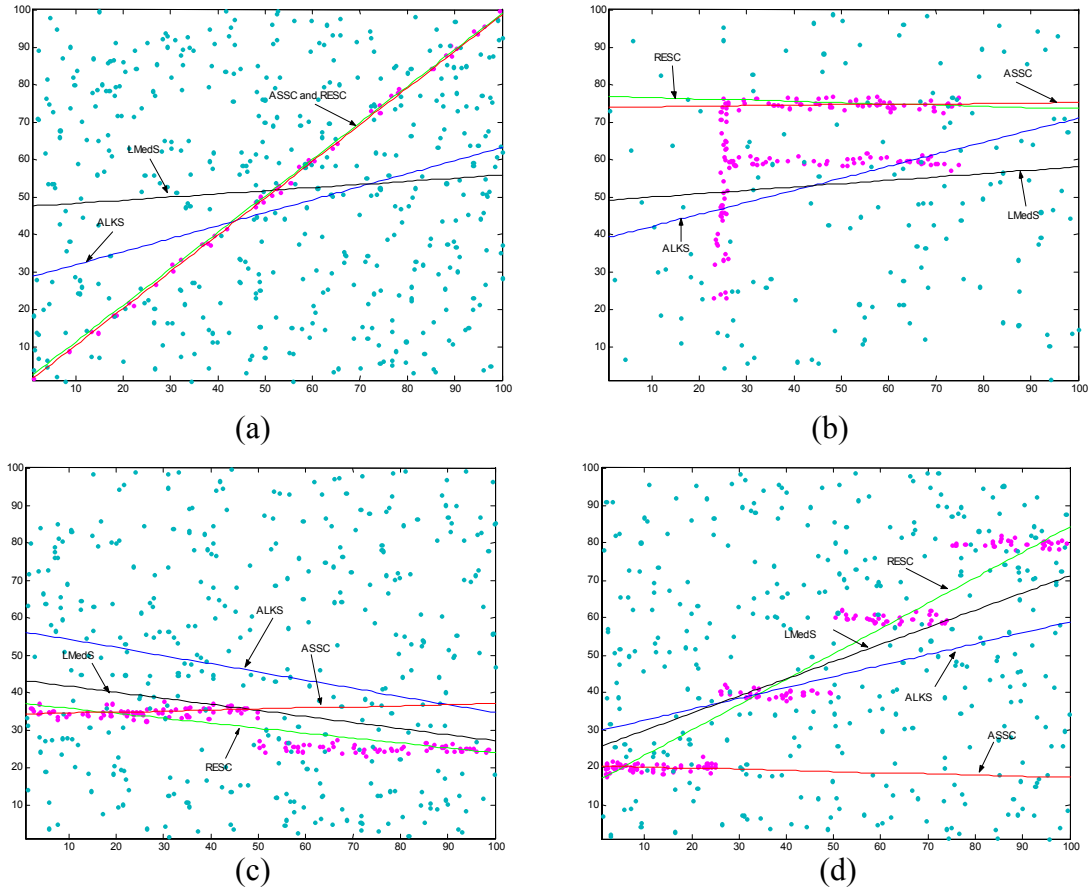

Fig. 5. Comparing the performance of four methods: (a) fitting a line with a total of 90% outliers; (b) fitting three lines with a total of 88% outliers; (c) fitting a step with a total of 85% outliers; (d) fitting three steps with a total of 89% outliers.

We generated four kinds of data (a line, three lines, a step, and three steps), each with a total of 500 data points. The signals were corrupted by Gaussian noise with zero mean and standard variance σ. Among the 500 data points, α data points were randomly distributed in the range of (0, 100). The i'th structure has γ_i data points.

(a) One line: x:(0-100), y=x, γ_1=50; α=450; σ=0.8.
(b) Three lines: x:(25-75), y=75, γ_1=60; x:(25-75), y=60, γ_2=50; x=25, y:(20-75), γ_3=40; α=350; σ=1.0.
(c) One step: x:(0-50), y=35, γ_1=75; x:(50-100), y=25, γ_2=55; α=370; σ=1.1.
(d) Three steps: x:(0-25), y=20, γ_1=55; x:(25-50), y=40, γ_2=30; x:(50-75), y=60, γ_3=30; x:(75-100), y=80, γ_4=30; α=355; σ=1.0.

From figure 5, we can see that the proposed ASSC method yields the best results among the four comparative methods and correctly fits all four signals. Because LMedS has only a 0.5 breakdown

point, it cannot resist the influence of more than 50% outliers. LMedS failed to fit all the four signals. Although ALKS can tolerate more than 50% outliers, the experimental results show that ALKS is not applicable for the signals with such large percentages of outliers because it failed in all four cases. RESC gave better results than LMedS and ALKS. It succeeded in two cases (one-line and three-line signals) even when the data involved more than 88% outliers. However, RESC failed to fit two signals (one step and three steps).

It should be emphasized that both the bandwidth choice and the scale estimation in the proposed method are data-driven. No priori knowledge about the bandwidth and the scale is necessary in the proposed method. This is a great improvement over the traditional RANSAC method where the user set a priori scale-related error bound.

6.2 3D examples

Two synthetic 3D signals were generated. Each contained 500 data points and three planar structures. Each plane contains 100 points corrupted by Gaussian noise with standard variance σ; 200 points are randomly distributed in a region including all three structures. A planar equation can be written as $Z=AX+BY+C$, and the residual of the point at (X_i, Y_i, Z_i) is $r_i=Z_i-AX_i-BY_i-C$. $(A, B, C; \sigma)$ are the parameters and scale to estimate.

In order to extract all planes, (1) we apply the robust estimators to the data set and estimate the parameters and scale of a plane; (2) we extract the inliers and remove them from the data set; (3) we repeat step 1 to 2 until all planes are extracted. The red circle is the first plane extracted; the green star is the second plane extracted; and the blue square is the third extracted plane. The results are shown in figure 6, table 1; figure 7 and table 2 (due to the limit of space, the results of LMedS, which completely broke down for these 3D data, are only given in table 1 and 2).

From figure 6 and table 1, we can see that RESC and ALKS, which claim to be robust to data with more than 50% outliers, fit the first plane approximately correctly. However, because the estimated scales for the first plane are totally wrong, these two methods failed to fit the second and third planes. Because the LMedS has only a 50% breakdown point, it completely failed to fit data with such high contamination — 80% outliers (see table 1). As a comparison, the proposed method yielded the best results. It successfully fitted all three planes and correctly estimates the scales of the three planes (the extracted three planes by the proposed method are shown in figure 6 (b)).

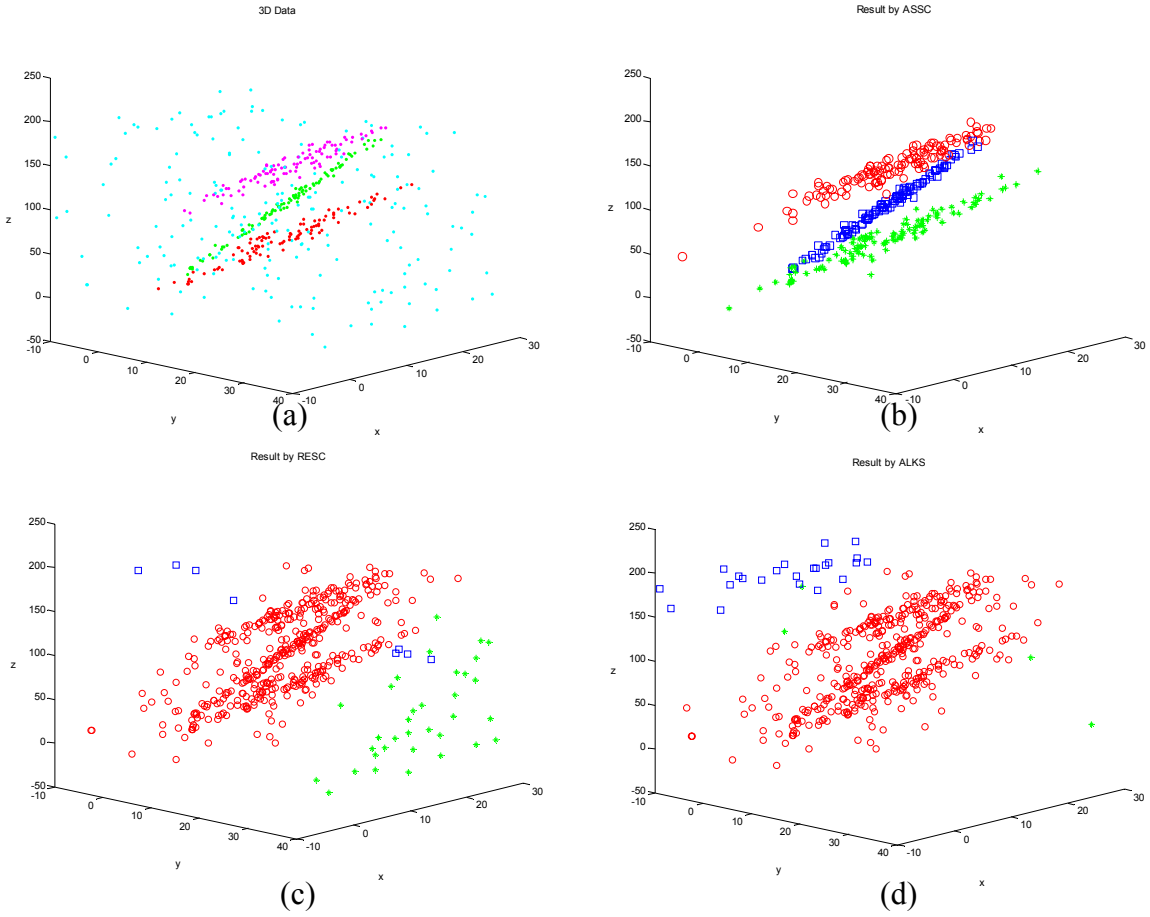

Fig. 6. First experiment for 3D multiple-structure data: (a) the 3D data; the results by (b) the proposed method; (c) RESC; (d) ALKS.

	Plane A	Plane B	Plane C
True values	(3.0, 5.0, 0.0; 3.0)	(2.0, 3.0, 0.0; 3.0)	(2.0, 3.0, 80.0; 3.0)
ASSC	(3.02, 4.86, 1.66; 3.14)	(2.09, 2.99, 0.56, 3.18)	(1.79, 2.98, 83.25, 3.78)
RESC	(3.69, 5.20, -7.94, 36.94)	(4.89, 13.82, -528.06,51.62) and (-2.88,-1.48, 189.65,0.47)	
ALKS	(2.74, 5.08, 1.63; 44.37)	(-7.20, 0.91,198.10; 0.007) and (-0.59,1.82,194.06; 14.34)	
LMedS	(1.22, 3.50,30.36, 51.50), (-0.11, -3.98, 142.80; 31.31) and (-9.59, -1.66,251.24;0.0)		

Table 1. Result comparisons of the four robust estimators for data in figure 6.

Similarly, in the second experiment (figure 7 and table 2), LMedS and ALKS completely broke down for the heavily corrupted data with multiple structures. RESC, although it correctly fitted the first plane, wrongly estimated the scale the plane. RESC wrongly fitted the second and the third planes. Only the proposed method correctly fitted all three planes (figure 7 (b)) and estimated the corresponding scale for each plane.

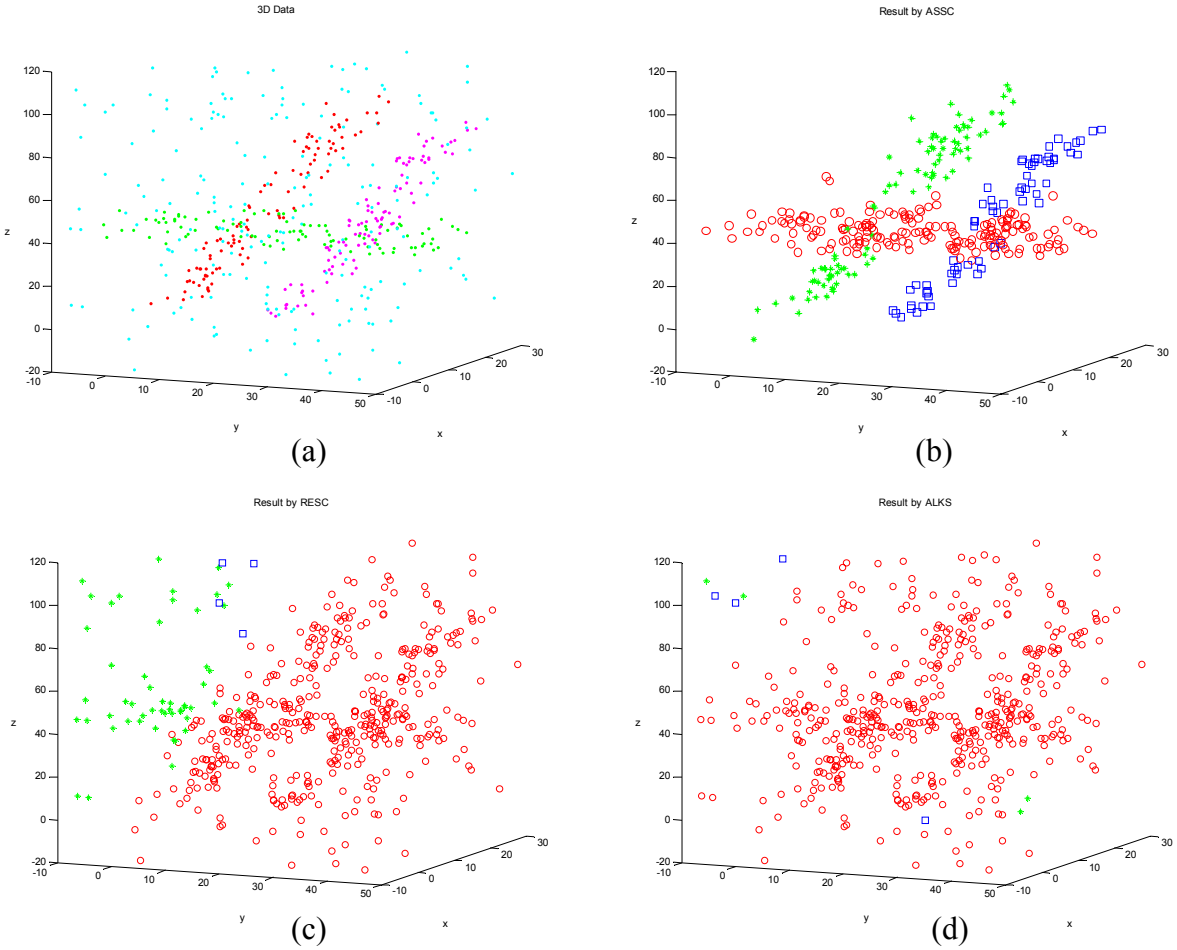

Fig. 7. Second experiment for 3D multiple-structure data: (a) the 3D data; the results by (b) the proposed method; (c) RESC; (d) ALKS.

	Plane A	Plane B	Plane C
True values	(0.0, 3.0, -60.0; 3.0)	(0.0, 3.0, 0.0; 3.0)	(0.0, 0.0, 40.0; 3.0)
ASSC	(0.00, 2.98, -60.68, 2.11)	(0.18, 2.93, 0.18, 3.90)	(0.08, 0.03, 38.26; 3.88)
RESC	(0.51, 3.04,-67.29;36.40)	(6.02,-34.00,-197.51;101.1) and (0.35, -3.85, 122.91, 0.02)	
ALKS	(-1.29, 1.03,14.35; 30.05), (-1.07, -2.07,84.31; 0.01) and (1.85, -11.19, 36.97; 0.08)		
LMedS	(0.25, 0.61,24.50, 27.06), (-0.04, -0.19, 92.27; 9.52) and (-0.12, -0.60,92.19; 6.89)		

Table 2. Result comparisons of the four robust estimators for data in figure 7.

The proposed method is computationally efficient. We perform the proposed method in MATLAB code with TSSE in Mex. When m is set as 500, the proposed method takes about 1.5 second for the 2D examples and about 2.5 seconds for the 3D examples in an AMD 800MHz personal computer.

7. Conclusions

This paper has shown that the scale estimation for data involving multiple structures and high percentages of outliers is as yet a relatively unsolved problem and more work in this field is required. This provides an important warning to the computer vision community: it is necessary to carefully choose a proper scale estimator.

A robust two-step scale estimator (TSSE) is proposed and comparative experiments showing its advantages over other existing robust scale estimators are given in this paper. TSSE can be used to give an initial scale estimate for robust estimators such as M-estimators, etc. TSSE can also be used as an auxiliary estimate of scale (after the parameters of a model to fit have been found) by other robust fitting methods such as Hough Transform [13], MDPE [26], etc.

Moreover, we, based on TSSE and RANSAC, propose a very robust Adaptive Scale Sample Consensus (ASSC) estimator. The proposed ASSC method considers both the number of data points within an error bound and the corresponding scale in its objective function. No priori knowledge about the scale of inliers is necessary in the proposed method. The proposed ASSC method can output the parameters of a model and the corresponding scale as its results. It is very robust to multiple-structural data containing high percentages of outliers. The computational cost of the proposed method is also low, which makes it applicable in many computer vision tasks. In [28], we will apply the proposed ASSC estimator to a more complicated computer vision task: range image segmentation.

References:

1. Bab-Hadiashar, A. and D. Suter, *Robust segmentation of visual data using ranked unbiased scale estimate.* ROBOTICA, International Journal of Information, Education and Research in Robotics and Artificial Intelligence, 1999. **17**: pp. 649-660.
2. Chen, H. and P. Meer, *Robust Computer Vision through Kernel Density Estimation.* A. Heyden et al. (Eds.), ECCV 2002, LNCS 2350, Springer-Verlag Berlin, 2002: pp. 236-250.
3. Chen, H., P. Meer, and D.E. Tyler. *Robust Regression for Data with Multiple Structures*. in *Proceedings of 2001 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2001. Kauai, Hawaii. pp. 1069-1075
4. Cheng, Y., *Mean Shift, Mode Seeking, and Clustering.* IEEE Trans. Pattern Analysis and Machine Intelligence, 1995. **17**(8): pp. 790-799.
5. Comaniciu, D. and P. Meer, *Robust Analysis of Feature Spaces: Color Image Segmentation.* in Proceedings of 1997 IEEE Conference on Computer Vision and Pattern Recognition, San Juan, PR, 1997: pp. 750-755.

6. Comaniciu, D. and P. Meer, *Mean Shift Analysis and Applications.* in Proceedings 7th International Conference on Computer Vision, Kerkyra, Greece, 1999: pp. 1197-1203.
7. Comaniciu, D. and P. Meer, *Mean Shift: A Robust Approach towards Feature Space A Analysis.* IEEE Trans. Pattern Analysis and Machine Intelligence, 2002. **24**(5): pp. 603-619.
8. Comaniciu, D., V. Ramesh, and A.D. Bue. *Multivariate Saddle Point Detection for Statistical Clustering*. in *Europe Conference on COmputer Vision (ECCV)*. 2002. Copenhagen, Danmark. pp. 561-576
9. Comaniciu, D., V. Ramesh, and P. Meer. *The Variable Bandwidth Mean Shift and Data-Driven Scale Selection*. in *Proc. Eighth Int'l Conf. Computer Vision*. 2001. pp. 438-445
10. Fischler, M.A. and R.C. Rolles, *Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography.* Commun. ACM, 1981. **24**(6): pp. 381-395.
11. Fukunaga, K. and L.D. Hostetler, *The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition.* IEEE Trans. Info. Theory, 1975. **IT-21**: pp. 32-40.
12. Haralick, R.M., *Computer vision theory: The lack thereof.* Comput. Vision Graphics Image Processing, 1986. **36**: pp. 372-386.
13. Hough, P.V.C., *Methods and means for recognising complex patterns*, in *U.S. Patent 3 069 654*. 1962.
14. Huber, P.J., *Robust Statistics.* New York, Wiley, 1981.
15. Leavers, V.F., *Which Hough transform?* CVGIP: Image Understanding, 1993. **58**(2): pp. 250-264.
16. Lee, K.-M., P. Meer, and R.-H. Park, *Robust Adaptive Segmentation of Range Images.* IEEE Trans. Pattern Analysis and Machine Intelligence, 1998. **20**(2): pp. 200-205.
17. Miller, J.V. and C.V. Stewart, *MUSE: Robust Surface Fitting Using Unbiased Scale Estimates.* Proc. Computer Vision and Pattern Recognition, San Francisco, 1996. **300-306**.
18. Rousseeuw, P.J., *Least Median of Squares Regression.* J. Amer. Stat. Assoc, 1984. **79**: pp. 871-880.
19. Rousseeuw, P.J. and C. Croux, *Alternatives to the Median Absolute Deriation.* Journal of the American Statistical Associaion, 1993. **88**(424): pp. 1273-1283.
20. Rousseeuw, P.J. and A. Leroy, *Robust Regression and outlier detection.* John Wiley & Sons, New York., 1987.
21. Siegel, A.F., *Robust Regression using Repeated Medians.* Biometrika, 1982. **69**: pp. 242-244.
22. Silverman, B.W., *Density Estimation for Statistics and Data Analysis*. 1986, London: Chapman and Hall.
23. Stewart, C.V., *MINPRAN: A New Robust Estimator for Computer Vision.* IEEE Trans. Pattern Analysis and Machine Intelligence, 1995. **17**(10): pp. 925-938.
24. Stewart, C.V., *Bias in Robust Estimation Caused by Discontinuities and Multiple Structures.* IEEE Trans. Pattern Analysis and Machine Intelligence, 1997. **19**(8): pp. 818-833.
25. Wand, M.P. and M. Jones, *Kernel Smoothing.* Chapman & Hall, 1995.
26. Wang, H. and D. Suter, *MDPE: A Very Robust Estimator for Model Fitting and Range Image Segmentation.* Submitted to IJCV, 2002.
27. Wang, H. and D. Suter. *False-Peaks-Avoiding Mean Shift Method for Unsupervised Peak-Valley Sliding Image Segmentation.* MECSE-2003-1: Monash University.
28. Wang, H. and D. Suter. *A Model-Based Range Image Segmentation Algorithm Using a Novel Robust Estimator*. in *Submitted to 3rd Internation Workshop on Statistical and Computational Theories of Vision (SCTV)*. 2003
29. Yu, X., T.D. Bui, and A. Krzyzak, *Robust Estimation for Range Image Segmentation and Reconstruction.* IEEE Trans. Pattern Analysis and Machine Intelligence, 1994. **16**(5): pp. 530-538.