Robust Adaptive-Scale Parametric Model Estimation for Computer Vision

Hanzi Wang and David Suter, *Senior Member*, *IEEE*

Abstract—Robust model fitting essentially requires the application of two estimators. The first is an estimator for the values of the model parameters. The second is an estimator for the scale of the noise in the (inlier) data. Indeed, we propose two novel robust techniques: the Two-Step Scale estimator (TSSE) and the Adaptive Scale Sample Consensus (ASSC) estimator. TSSE applies nonparametric density estimation and density gradient estimation techniques, to robustly estimate the scale of the inliers. The ASSC estimator combines Random Sample Consensus (RANSAC) and TSSE: using a modified objective function that depends upon both the number of inliers and the corresponding scale. ASSC is very robust to discontinuous signals and data with multiple structures, being able to tolerate more than 80 percent outliers. The main advantage of ASSC over RANSAC is that prior knowledge about the scale of inliers is not needed. ASSC can simultaneously estimate the parameters of a model and the scale of the inliers belonging to that model. Experiments on synthetic data show that ASSC has better robustness to heavily corrupted data than Least Median Squares (LMedS), Residual Consensus (RESC), and Adaptive Least Kth order Squares (ALKS). We also apply ASSC to two fundamental computer vision tasks: range image segmentation and robust fundamental matrix estimation. Experiments show very promising results.

Index Terms—Robust model fitting, random sample consensus, least-median-of-squares, residual consensus, adaptive least kth order squares, kernel density estimation, mean shift, range image segmentation, fundamental matrix estimation.

1 Introduction

Robust parametric model estimation techniques have been used with increasing frequency in many computer vision tasks such as optical flow calculation [1], [22], [38], range image segmentation [15], [19], [20], [36], [39], estimating the fundamental matrix [33], [34], [40], and tracking [3]. A robust estimation technique is a method that can estimate the parameters of a model from inliers and resist the influence of outliers. Roughly, outliers can be classified into two classes: gross outliers and pseudo-outliers [30]. Pseudo-outliers contain structural information, i.e., pseudo-outliers can be outliers to one structure of interest but inliers to another structure. Ideally, a robust estimation technique should be able to tolerate both types of outliers. Multiple structures occur in most computer vision problems. Estimating information from data with multiple structures remains a challenging task despite the search for highly robust estimators in recent decades [4], [6], [20], [29], [36], [39]. 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 ([25, p. 9]).

Although the least squares (LS) method can achieve optimal results under Gaussian distributed noise, only one single outlier is sufficient to force the LS estimator to produce an arbitrarily large value. Thus, robust estimators have been proposed in the statistics literature [18], [23], [25], [27] and in the computer vision literature [13], [17], [20], [29], [36], [38], [39].

• *The authors are with the Department of Electrical and Computer Systems Engineering, Monash University, Clayton Vic. 3800, Australia. E-mail: {hanzi.wang, d.suter}@eng.monash.edu.au.*

Manuscript received 28 July 2003; revised 31 Mar. 2004; accepted 5 Apr. 2004.
Recommended for acceptance by Y. Amit.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0196-0703.

Traditional statistical estimators have breakdown points that are no more than 50 percent. These robust estimators assume that the inliers occupy the absolute majority of the whole data, which is far from being satisfied for the real tasks faced in computer vision [36]. It frequently happens that outliers occupy the absolute majority of the data. Although Rousseeuw and Leroy argue that 0.5 is the theoretical maximum breakdown point [25], the proof shows that they require the robust estimator has a unique solution, (more technically, they require affine equivariance) [39]. As Stewart noted [31], [29], a breakdown point of 0.5 can and must be surpassed.

A number of recent estimators claim to have a tolerance to more than 50 percent outliers. Included in this category of estimators, although no formal proof of high breakdown point exists, are the Hough transform [17] and the RANSAC method [13]. However, they need a user to set certain parameters that essentially relate to the level of noise expected: a priori an estimate of the scale, which is not available in most practical tasks. If the scale is wrongly provided, these methods will fail.

RESC [39], MINPRAN [29], MUSE [21], ALKS [20], MSSE [2], etc., all claim to be able to tolerate more than 50 percent outliers. However, RESC needs the user to tune many parameters in compressing a histogram. MINPRAN assumes that the outliers are randomly distributed within a certain range, which makes MINPRAN less effective in extracting multiple structures. Another problem of MINPRAN is its high computational cost. MUSE and ALKS are limited in their ability to handle extreme outliers. MUSE also needs a lookup table for the scale estimator correction. Although MSSE can handle large percentages of outliers and pseudo-outliers, it does not seem as successful in tolerating extreme cases.

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 (mean shift).
- By employing TSSE in a RANSAC like procedure, we propose a highly robust estimator: Adaptive Scale Sample Consensus (ASSC) estimator. ASSC is an important improvement over RANSAC because no priori knowledge concerning the scale of inliers is necessary (the scale estimation is data driven). Empirically, ASSC can tolerate more than 80 percent outliers.
- Experiments presented show that both TSSE and ASSC are highly robust to heavily corrupted data with multiple structures and discontinuities, and that they outperform several competing methods. These experiments also include real data from two important tasks: range image segmentation and fundamental matrix estimation.

This paper is organized as follows: In Section 2, we review previous robust scale techniques. In Section 3, density gradient estimation and the mean shift/mean shift valley method are introduced, and a 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 experimental comparisons, using both 2D and 3D examples, are contained in Section 6. We apply ASSC to range image segmentation in Section 7 and fundamental matrix estimation in Section 8. We conclude in Section 9.

2 Robust Scale Estimators

Differentiating outliers from inliers usually depends crucially upon whether the scale of the inliers has been correctly estimated. Some robust estimators, such as 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, etc., produce a robust estimate of scale (after finding the parameters of a model) during a postprocessing stage, which aims to differentiate inliers from outliers. Recent work of Chen and Meer [4], [5] sidesteps scale estimation per-se by deciding the inlier/outlier threshold based upon the valleys either side of the mode of projected residuals (projected on the direction normal to the hyperplane of best fit). This has some similarity to our approach in that they also use Kernel-Density estimators and peak/valley seeking on that kernel density estimate (peak by mean shift, as we do; valley by a form of search as opposed to our mean shift valley method). However, their method is not a direct attempt to estimate scale nor is it as general as the approach here (we are not restricted to finding linear/hyperplane fits). Moreover, we do not have a (potentially) costly search for the normal direction that maximizes the concentration of mass about the mode of the kernel density estimate as in Chen and Meer.

Given a scale estimate, s, the inliers are usually taken to be those data points that satisfy the following condition:

$$|r_i/s| < T, \tag{1}$$

where r_i is the residual of ith sample, and T is a threshold. For example, if T is 2.5 (1.96), 98 percent (95 percent) of a Gaussian distribution will be identified as inliers.

2.1 The Median and Median Absolute Deviation (MAD) Scale Estimator

Among many robust scale estimators, the sample median is popular. The sample median is bounded when the data include more than 50 percent inliers. A robust median scale estimator is then given by [25]:

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

where n is the number of sample points and p is the dimension of the parameter space (e.g., two for a line, three for a circle).

A variant, MAD, which recognizes that the data points may not be centered, uses the median to center the data [24]:

$$MAD = 1.4826 med_i\{|r_i - med_j r_j|\}. \tag{3}$$

The median and MAD estimators have breakdown points of 50 percent. Moreover, both methods are biased for multiple-mode cases even when the data contains less than 50 percent outliers (see Section 4).

2.2 Adaptive Least Kth Squares (ALKS) Estimator

A generalization of median and MAD (which both use the median statistic) is to use the kth order statistic in ALKS [20]. This robust k scale estimate, assuming inliers have a Gaussian distribution, is given by:

$$\hat{s}_k = \frac{\hat{d}_k}{\Phi^{-1}[(1+k/n)/2]}, \tag{4}$$

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 claimed [20] to be that which corresponds to the minimum of the variance of the normalized error ε_k^2:

$$\varepsilon_k^2 = \frac{1}{k-p}\sum_{i=1}^{k}\left(\frac{r_{i,k}}{\hat{s}_k}\right)^2 = \frac{\hat{\sigma}_k^2}{\hat{s}_k^2}. \tag{5}$$

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

2.3 Modified Selective Statistical Estimator (MSSE)

Bab-Hadiashar and Suter [2] also use the least kth order (rather than median) residuals and have a heuristic way of determining inliers that relies on finding the last "reliable" unbiased scale estimate as residuals of larger and larger value are included. That is, after finding a candidate fit to the data, they try to recognize the first outlier, corresponding to where the kth residual "jumps," by looking for a jump in the unbiased scale estimate formed by using the first kth residuals in an ascending order:

$$\hat{\sigma}_k^2 = \frac{\sum_{i=1}^{k} r_i^2}{k-p}. \tag{6}$$

Essentially, the emphasis is shifted from using a good scale estimate for defining the outliers, to finding the point of breakdown in the unbiased scale estimate (thereby signaling the inclusion of an outlier). This breakdown is signaled by the first k that satisfies the following inequality:

$$\frac{\sigma_{k+1}^2}{\sigma_k^2} > 1 + \frac{T^2 - 1}{k - p + 1}. \tag{7}$$

2.4 Residual Consensus (RESC) Method

In RESC [39], after finding a fit, one estimates the scale of the inliers by directly calculating:

$$\sigma = \alpha \left(\frac{1}{\sum_{i=1}^{v} h_i^c - 1} \sum_{i=1}^{v} (i h_i^c \delta - \bar{h}^c)^2 \right)^{1/2}, \tag{8}$$

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

However, we find that the estimated scale is overestimated because, instead of summing up the squares of the differences between all individual residuals and the mean residual in the CH, (8) sums up the squares of the differences between residuals in each bin of CH and the mean residual in the CH.

To reduce this problem, we propose an alternative form:

$$\sigma = \left(\frac{1}{\sum_{i=1}^{v} h_i^c - 1} \sum_{i=1}^{n_c} (r_i - \bar{h}^c)^2 \right)^{1/2}, \tag{9}$$

where n_c is the number of data points in the CH. We compare our proposed improvement in experiments reported later in this paper.

3 A Robust Scale Estimator: TSSE

In this section, we propose a highly robust scale estimator (TSSE), which is derived from kernel density estimation techniques and the mean shift method. We review these foundations first.

3.1 Density Gradient Estimation and Mean Shift Method

When one has *samples* drawn from a distribution, there are several nonparametric methods available for estimating that density of those samples: the histogram method, the naive method, the nearest neighbour method, and kernel density estimation [28].

The multivariate kernel density estimator with kernel K and window radius (bandwidth) h is defined as follows ([28, p. 76]):

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h} \right) \tag{10}$$

for a set of n data points $\{\mathrm{X_i}\}\mathrm{i} = 1, \ldots, \mathrm{n}$ in a d-dimensional Euclidian space R^d and $K(\mathrm{x})$ satisfying some conditions ([35, p. 95]). The Epanechnikov kernel ([28, p. 76]) is often used as it yields the minimum mean integrated square error (MISE):

$$K_e(\mathrm{X}) = \begin{cases} \frac{1}{2} c_d^{-1} (d+2)(1 - \mathrm{X}^T \mathrm{X}) & if\ \mathrm{X}^T \mathrm{X} < 1 \\ 0 & otherwise, \end{cases} \tag{11}$$

where c_d is the volume of the unit d-dimensional sphere, e.g., $c_1 = 2,\ c_2 = \pi,\ c_3 = 4\pi/3$. (Note: there are other possible criteria for optimality, suggesting alternative kernels—an issue we will not explore here.)

To estimate the gradient of this density, we can take the gradient of the kernel density estimate

$$\hat{\nabla} f(x) \equiv \nabla \hat{f}(x) = \frac{1}{nh_d} \sum_{i=1}^{n} \nabla K\left(\frac{x - X_i}{h} \right). \tag{12}$$

According to (12), the density gradient estimate of the Epanechnikov kernel can be written as

$$\hat{\nabla} f(x) = \frac{n_x}{n(h^d c_d)} \frac{d+2}{h^2} \left(\frac{1}{n_x} \sum_{X_i \in S_h(x)} [X_i - x] \right), \tag{13}$$

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

It is useful to define the mean shift vector $\mathrm{M_h(x)}$ [14] as:

$$\mathrm{M_h(x)} \equiv \frac{1}{n_x} \sum_{X_i \in S_h(x)} [X_i - x] = \frac{1}{n_x} \sum_{X_i \in S_h(x)} X_i - x. \tag{14}$$

Thus, (13) can be rewritten as:

$$\mathrm{M_h(x)} \equiv \frac{h^2}{d+2} \frac{\hat{\nabla} f(x)}{\hat{f}(x)}. \tag{15}$$

Fukunaga and Hostetler [14] observed that the mean shift vector points toward the direction of the maximum increase in the density, thereby suggesting a mean shift method for locating the peak of a density distribution. This idea has recently been extensively exploited in low level computer vision tasks [7], [9], [10], [8].

3.2 Mean Shift Valley Algorithm

Sometimes it is very important to find the valleys of distributions. Based upon the Gaussian kernel, a saddle-point seeking method was published in [11]. Here, we describe a more simple method, based upon the Epanechnikov kernel [37]. We define the mean shift valley vector to point in the opposite direction to the peak:

$$\mathrm{MV_h}(x) = -\mathrm{M_h}(x) = x - \frac{1}{n_x} \sum_{X_i \in S_h(x)} X_i. \tag{16}$$

In practice, we find that the step-size given by the above equation may lead to oscillation. Let $\{y_k\}_{k=1,2\ldots}$ be the sequence of successive locations of the mean shift valley procedure, then we take a modified step by:

$$\mathrm{y_{k+1}} = \mathrm{y_k} + p \cdot MV_h(y_k), \tag{17}$$

where $0 < p \leq 1$. To avoid oscillation, we adjust p so that $\mathrm{MV_h(y_k)^T MV_h(y_{k+1})} > 0$.

Note: when there are no local valleys (e.g., unimode), the mean shift valley method is divergent. This can be easily detected and avoided by terminating when no data samples fall within the window.

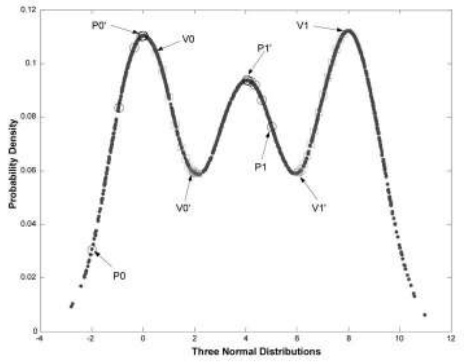

Fig. 1. An example of applying the mean shift method to find local peaks and applying the mean shift valley method to find local valleys.

3.3 Bandwidth Choice

One crucial issue in nonparametric density estimation and, thereby, in the mean shift method, and in the mean shift valley method, is how to choose the bandwidth h [10], [12], [35]. Since we work in one-dimensional residual space, a simple over-smoothed bandwidth selector is employed [32]:

$$\hat{h} = \left[\frac{243R(K)}{35u_2(K)^2 n}\right]^{1/5} S, \tag{18}$$

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, the MAD, or the 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 integrated squared error) optimal bandwidth $\hat{h}_{AMISE}$. The median, MAD, and robust scale estimator may be biased for data with multimodes. This is because these estimators are proposed assuming the whole data have a Gaussian distribution. Because the bandwidth in (18) is proportional to the estimated scale, the bandwidth can be set as $c\,\hat{h}, (0 < c < 1)$ to avoid oversmoothing ([35], p. 62).

To illustrate, we generated the data in Fig. 1. Mode 1 has 600 data points (mean 0.0), mode 2 has 500 data points (mean 4.0), and mode 3 has 600 data points (mean 8.0). We set two initial points: P0 (-2.0) and P1(5.0) and, after applying the mean shift method, we obtained the two local peaks: P0'(0.01) and P1'(4.03). Similarly, we applied the mean shift valley method to two selected two initial points: V0 (0.5) and V1 (7.8). The valley V0' was located at 2.13, and V1' was located at 6.00.

3.4 Two-Step Scale Estimator (TSSE)

Based upon the above mean-shift based procedures, we propose a robust two-step method to estimate the scale of the inliers.

1. Use mean shift, with initial center zero (in ordered absolute residual space), to find the local peak, and then use the mean shift valley to find the valley next to the peak. Note: modes other than the inliers will be disregarded as they lie outside the obtained valley.
2. Estimate the scale of the fit by the median scale estimator using the points within the band centered at the local peak extending to the valley.

TSSE is very robust to outliers and can resist heavily contaminated data with multiple structures. In the next section, we will compare this method with five other methods.

4 Experimental Comparisons of Scale Estimation

In this section, we investigate the behavior of several robust scale estimators that are widely used in computer vision community: showing some of the weaknesses of these scale estimation techniques.

4.1 Experiments on Scale Estimation

Three experiments are included here (see Table 1). In experiment 1, there was only one structure in the data, in experiments 2 and 3, there were two structures in the data. In the description, the ith structure has $\mathrm{n_i}$ data points, all corrupted by Gaussian noise with zero mean and standard variance σ_i, and α outlier data points were randomly distributed in the range of (0, 100). For the purposes of this section (only), we *assume we know the parameters of the model*: this is so we can concentrate on estimating the scale of the residuals. In experiments 2 and 3, we assume we know the parameters of the first structure (highlighted in bold) and it is the parameters of that structure we use for the calculation of the residuals. The second structure then provides pseudo-outliers to the first structure.

We note that even when there are no outliers (experiment 1) ALKS performs poorly. 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 (5) estimates the optimal k wrongly. ALKS classified only 10 percent of the data as inliers. (Note: since the TSSE use the median of the inliers to define the scale, it is no surprise that, when 100 percent of the data are inliers, it produces the same estimate as the median.)

From the obtained results, we can see that only the proposed method gave a reasonably good result when the number of outliers is extreme (experiment 3).

4.2 Error Plots

We also generated two signals for the error plots (similar to the breakdown plots in [25]):

- **Roof Signal**—500 data points in total: x: (0-55), $\mathrm{y}=\mathrm{x}+30$, n_1, $\sigma = 2$; x:(55-100), $\mathrm{y} = 140 - \mathrm{x}$, $n_2 = 50$; $\sigma = 2$. Where we first assigned 450 data point to n_1 and the number of the uniform outliers $\alpha = 0$. Thus, the data include 10 percent outliers. For each successive run, we decrease n_1, and at the same time, we increase α so that the total number of data points is 500. Finally, $n_1 = 75$ and $\alpha = 375$, i.e., the data include 85 percent outliers. The results are averaged over runs repeated 20 times at the same setting.
- **One-Step Signal**—1,000 data points in total: x:(0-55), $\mathrm{y} = 30$, n_1, $\sigma = 2$; x:(55-100), $\mathrm{y} = 40$, $n_2 = 100$; $\sigma = 2$. First, we assign n_1 900 data points and the number of the uniform outliers $\alpha = 0$. Thus, the data include 10 percent outliers. Then, we decrease n_1, and at the same time, we increase α so that the number of the whole data points is 1,000. Finally, $n_1 = 150$ and $\alpha = 750$, i.e., the data include 85 percent outliers.

Fig. 2 shows that TSSE yielded the best results among the six methods (it begins to break down only when outliers

TABLE 1
Comparison of Scale Estimation for Three Situations of Increasing Difficulty

Data Description	True Scale	Median	MAD	ALKS	MSSE	RESC (revised)	TSSE
1. Normal distribution: single mode One line: x:(0-55), y=30, n_1=500, σ_1=3; α=0, i.e., 100% inliers.	3.0	2.9531	2.9461	2.0104	2.6832	2.8143	2.9531
2. Two modes: step signal composed of: **line1: x:(0-55), y=40, n_1=3000, σ_1=3**; line2: x:(55-100), y=70, n_2=2000, σ_2=3; α=0 i.e. 40% pseudo outliers.	3.0	6.3541	8.8231	3.2129	2.8679	2.9295	3.0791
3. Two modes: same step signal as above with 80% that are outliers (addition of extra uniformly distributed outliers to the pseudo outliers), i.e., n_1=1000; n_2=750; α=3250.	3.0	34.0962	29.7909	7.2586	27.4253	24.4297	4.1427

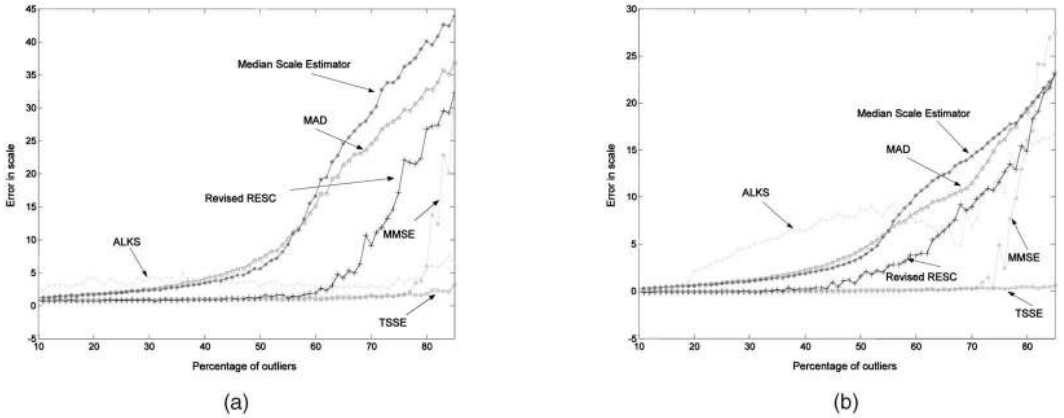

Fig. 2. Error plot of six methods in estimating the scale of a (a) roof and (b) step.

occupy more than 87 percent). The revised RESC method begins to break down when the outliers occupy around 60 percent for Fig. 2a and 50 percent for Fig. 2b. MSSE gave reasonable results when the percentage of outliers is less than 75 percent for Fig. 2a and 70 percent for Fig. 2b, but it broke down when the data include more outliers. ALKS yielded less accurate results than TSSE, and less accurate results than the revised RESC and MMSE when the outliers are less than 60 percent. Although the breakdown points of the median and the MAD scale estimators are as high as 50 percent, their results deviated from the true scale even when outliers are less than 50 percent of the data. They are biased more and more from the true scale with the increase in the percentage of outliers. Comparing Fig. 2a with Fig. 2b, we can see that the revised RESC, MSSE, and ALKS yielded less accurate results for a small step in the signal when compared to a roof signal, but the results of the proposed TSSE are of similar accuracy for both types of signals.

We also investigated the performance of the robust k scale estimator, for different choices of the "quantile" k, again assuming the correct parameters of a model have been found. Let:

$$S(q) = \frac{\hat{d}_q}{\Phi^{-1}[(1+q)/2]}, \tag{19}$$

where q is varied from 0 to 1. Note: S (0.5) is the median scale estimator.

We generated a one-step signal containing 500 data points in total: x:(0-55), $\mathrm{y} = 30$, n_1, $\sigma = 1$; x:(55-100), $\mathrm{y} = 40$, $n_2 = 50$; $\sigma = 1$. At the beginning, $n_1 = 450$ and $\alpha = 0$. Then, we decrease n_1, and at the same time, we increase α until $n_1 = 50$, and $\alpha = 400$, i.e., the data include 90 percent outliers.

As Fig. 3 shows, the accuracy of $S(q)$ is increased with the decrease of q. When the outliers are less than 50 percent of the whole data, the difference for different values of q is small. However, when the data include more than 50 percent outliers, the difference for various values of q is large.

4.3 Performance of TSSE

From the experiments in this section, we can see the proposed TSSE is a very robust scale estimator, achieving better results than the other five methods. However, we must acknowledge that the accuracy of TSSE is related to the accuracy of kernel density estimation. In particular, for very few data points, the kernel density estimates will be less accurate. For example, we repeat the experiment 1 (in Section 4.1) using different numbers of the data points n_1. For each n_1, we repeat the experiment 100 times. Thus, the mean (mn) and the stand variance (std) of the results can be obtained: $n_1 = 300$, $\mathrm{mn} = 3.0459$, $\mathrm{std} = 0.2128$; $n_1 = 200$, $\mathrm{mn} = 3.0851$, $\mathrm{std} = 0.2362$; $n_1 = 100$, $\mathrm{mn} = 3.1843$, $\mathrm{std} = 0.3527$; $\mathrm{n}_1 = 50$, $\mathrm{mn} = 3.2464$, $\mathrm{std} = 0.6216$. The achievement of ASSC decreases with the reduction in the number of data points. We note that this also happens to the other five methods.

In practice, one cannot directly estimate the scale: the parameters of a model also need to be estimated. 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.

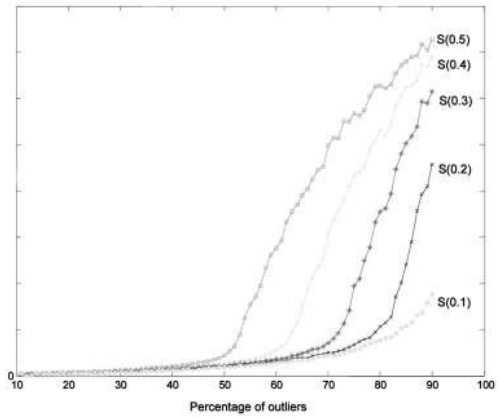

Fig. 3. Error plot of robust scale estimators based on different quantiles.

5 Robust Adaptive Scale Sample Consensus Estimator

Fischler and Bolles [13] introduced RANdom Sample Consensus (RANSAC). Like the common implementation of Least Median of Squares fitting, RANSAC randomly samples a p-subset (p is the dimension of parameter space) to estimate the parameters θ of the model. Also, like LMedS, RANSAC continues to draw randomly sampled p-subsets from the whole data until there is a high probability that at least one p-subset is clean. A p-tuple is "clean" if it consists of good observations without contamination by outliers. Let ε be the fraction of outliers contained in the whole set of points. The probability P, of one clean subset in m such subsets, can be expressed as follows ([25, p. 198]):

$$P = 1 - (1 - (1 - \varepsilon)^{\mathrm{p}})^{\mathrm{m}}. \tag{20}$$

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

$$m = \frac{\log(1 - P)}{\log[1 - (1-\varepsilon)^{p}]}. \tag{21}$$

The criterion that RANSAC uses to select the best fit is: Maximize the number of data points within a user-set error bound:

$$\hat{\theta} = \arg\max_{\hat{\theta}} n_{\hat{\theta}}, \tag{22}$$

where $n_{\hat{\theta}}$ is the number of points whose absolute residual is within the error bound.

The error bound in RANSAC is crucial. Provided with a correct error bound of inliers, the RANSAC method can find a model even when the data contain a large percentage of gross errors. However, when the error bound is wrongly given, RANSAC will totally break down even when the outliers occupy a small percentage of the whole data [36]. We use our scale estimator TSSE to automatically set the error bound—yielding a new parametric fitting scheme —ASSC, which includes both the number and the scale of inliers in its objective function.

5.1 Adaptive Scale Sample Consensus Estimator (ASSC) Algorithm

We assume that when a model is correctly found, two criteria 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.

We define our objective function as:

$$\hat{\theta} = \arg\max_{\hat{\theta}}(n_{\hat{\theta}}/S_{\hat{\theta}}). \tag{23}$$

Of course, there are many potential variants on this objective function but the above is a simple and natural one. Note: when the estimate of the scale is fixed, (23) is another form of RANSAC with the score n_θ scaled by $1/S$ (i.e., a fixed constant for all p-subsets), yielding the same results as RANSAC. ASSC is more reasonable because the scale is estimated for each candidate fit, in addition to the fact that it no longer requires a user defined error-bound.

The ASSC 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 the bandwidth by (18) and calculate an initial scale by a robust k scale estimator (19) using $q = 0.2$.
3. Apply TSSE 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 (10).
4. Validate the valley. Let $\hat{f}(valley)/\hat{f}(peak) = \lambda$ (where $0 \leq \lambda < 1$). Because the inliers are assumed to have 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 (21)). 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. In order to improve the statistical efficiency, a weighted least square procedure ([25, p. 202]) is 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 ASSC to obtain a high robustness that can tolerate more than 50 percent outliers, as demonstrated by the experiments in the next section.

6 Experiments with Data Containing Multiple Structures

In this section, both 2D and 3D examples are given. The results of the proposed method are also compared with those of three other popular methods: LMedS, RESC, and ALKS. All of these methods use the random sampling scheme that is also at the heart of our method. Note: unlike Section 4, *we do not, of course, assume any knowledge of the parameters of the models in the data. Nor are we aiming to find any particular structure. Due to the random sampling used, the methods will possibly return a different structure on different*

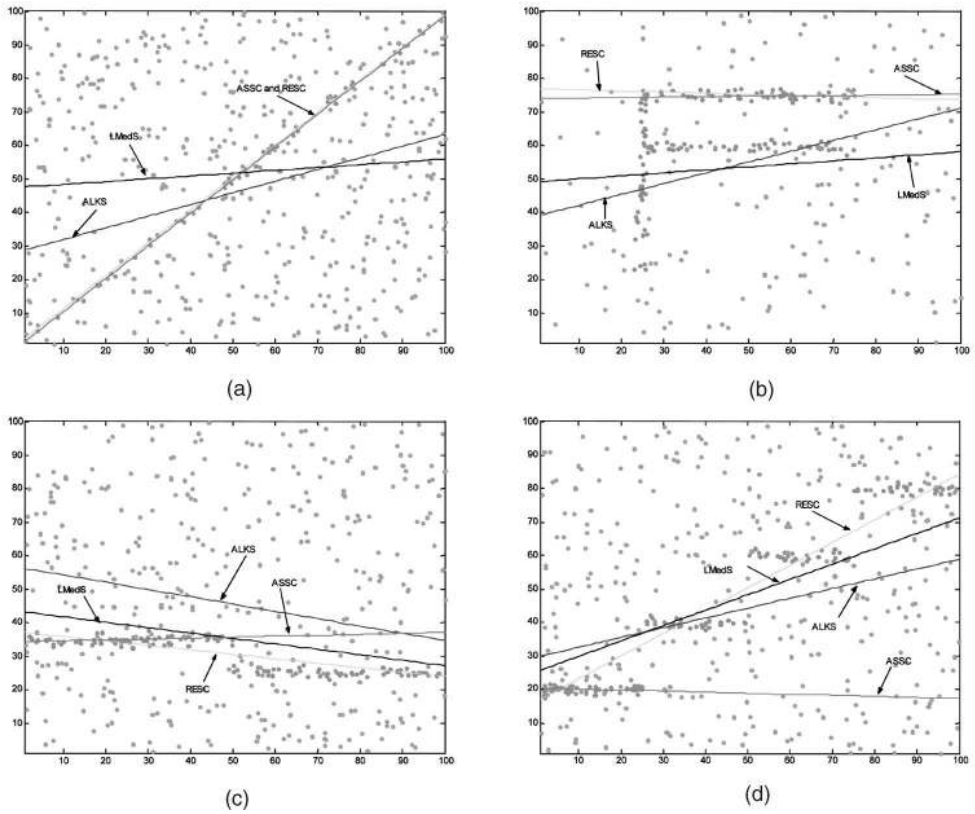

Fig. 4. Comparing the performance of four methods: (a) fitting a line with a total of 90 percent outliers, (b) fitting three lines with a total of 88 percent outliers, (c) fitting a step with a total of 85 percent outliers, (d) fitting three steps with a total of 89 percent outliers.

runs—however, they will generally find the largest structure most often, if one dominates in size.

6.1 2D Examples

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 $i\mathrm{th}$ structure has $\mathrm{n_i}$ data points.

1. One line: x:(0-100), $y = x$, $n_1 = 50$; $\alpha = 450$; $\sigma = 0.8$.
2. Three lines: x:(25-75), $y = 75$, $n_1 = 60$; x:(25-75), $y = 60$, $n_2 = 50$; $x = 25$, y:(20-75), $n_3 = 40$; $\alpha = 350$; $\sigma = 1.0$.
3. One step: x:(0-50), $\mathrm{y} = 35$, $n_1 = 75$; x:(50-100), $\mathrm{y} = 25$, $n_2 = 55$; $\alpha = 370$; $\sigma = 1.1$.
4. Three steps: x:(0-25), $\mathrm{y} = 20$, $n_1 = 55$; x:(25-50), $\mathrm{y} = 40$, $n_2 = 30$; x:(50-75), $\mathrm{y} = 60$, $n_3 = 30$; x:(75-100), $\mathrm{y} = 80$, $n_4 = 30$; $\alpha = 355$; $\sigma = 1.0$.

In Fig. 4, we can see that the proposed ASSC method yields the best results among the four methods, correctly fitting all four signals. Because LMedS has a 50 percent breakdown point, it failed to fit all the four signals. Although ALKS can tolerate more than 50 percent outliers, it failed in all four cases with very high outlier content. 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 percent outliers. However, RESC failed to fit two signals (Figs. 4c and 4d). (Note: If the number of steps in Fig. 4d increases greatly and each step gets short enough, ASSC, like others, cannot distinguish a series of very small steps from a single inclined line.)

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 must 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 $\mathrm{Z} = \mathrm{AX} + \mathrm{BY} + \mathrm{C}$, and the residual of the point at $(\mathrm{X_i}, \mathrm{Y_i}, \mathrm{Z_i})$ is $\mathrm{r_i} = \mathrm{Z_i} - \mathrm{AX_i} - \mathrm{BY_i} - \mathrm{C}$. $(\mathrm{A}, \mathrm{B}, \mathrm{C}; \sigma)$ are the parameters to estimate.

In contrast to the previous section, here we attempt to find all structures in the data. 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, and 3) we repeat Steps 1 to 2 until all planes are extracted. The red circles constitute the first plane extracted; the green stars the second plane extracted; and the blue squares the third extracted plane. The results are shown in Fig. 5, Table 2, Fig. 6, and Table 3 (due to the limited of space, the results of LMedS, which completely broke down for these 3D data, are only given in Tables 2 and 3). Note, for RESC, we use the revised form in (9) instead of (8) for scale estimate.

From Fig. 5 and Table 2, we can see that RESC and ALKS, which claim to be robust to data with more than 50 percent outliers, fit the first plane approximately correctly. However,

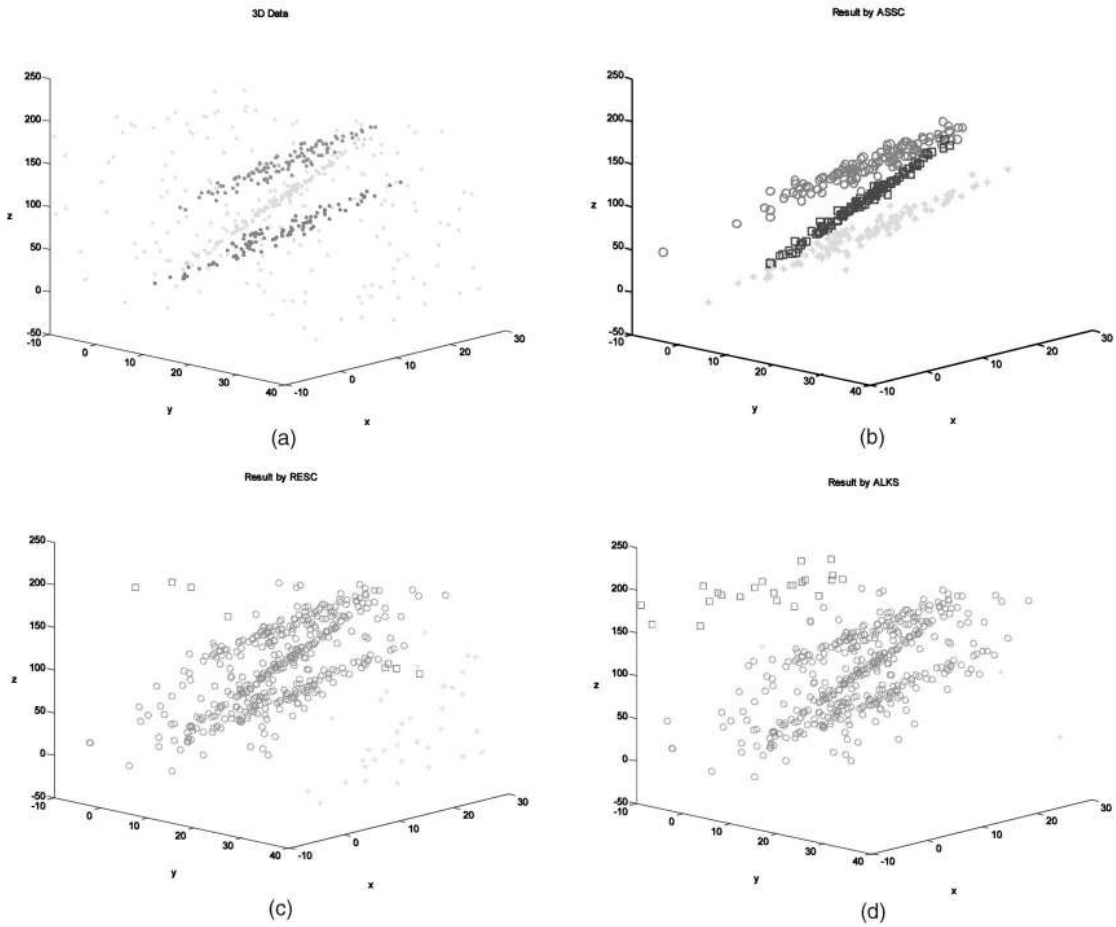

Fig. 5. First experiment for 3D multiple-structure data: (a) the 3D data; the results by (b) the proposed method, (c) by RESC, and (d) by ALKS.

TABLE 2
Result of the Estimates of the Parameters $(\mathrm{A}, \mathrm{B}, \mathrm{C}; \sigma)$ Provided by Each of the Robust Estimators Applied to the Data in Fig. 5

	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, 189.62,0.47)	
ALKS	(2.74, 5.08, 1.63; 44.37)	(-7.20, 0.91,198.1; 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)		

because the estimated scales for the first plane are quite wrong, these two methods failed to fit the second and third planes. LMedS, having a 50 percent breakdown point, completely failed to fit data with such high contamination (see Table 2). The proposed method yielded the best results: successfully fitting all three planes and correctly estimating the scales of the inliers to the three planes (the extracted three planes by the proposed method are shown in Fig. 5b).

Similarly, in the second experiment (Fig. 6 and Table 3), LMedS and ALKS completely broke down for the heavily corrupted data with multiple structures. RESC, although it correctly fitted the first plane, incorrectly estimated the scale of the inliers to the plane. RESC wrongly fitted the second and the third planes. Only the proposed method correctly fitted all three planes (Fig. 6b) and estimated the corresponding scale for each plane.

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

6.3 The Error Plot of the Four Methods

In this section, we perform an experiment to draw the error plot of each method (similar to the experiment reported in [39]. However, the data that we use is more complicated because it contains two types of outliers: clustered outliers and randomly distributed outliers). We generate one plane signal with Gaussian noise having unit standard variance. The clustered outliers have 100 data points and are distributed within a cube. The randomly distributed outliers contain the plane signal and clustered outliers. The number of inliers is decreased from 900 to 100. At the same time, the number of randomly distributed outliers is increased from 0 to 750 so that the total number of the data points is kept 1,000. Thus, the outliers occupy from 10 percent to 90 percent.

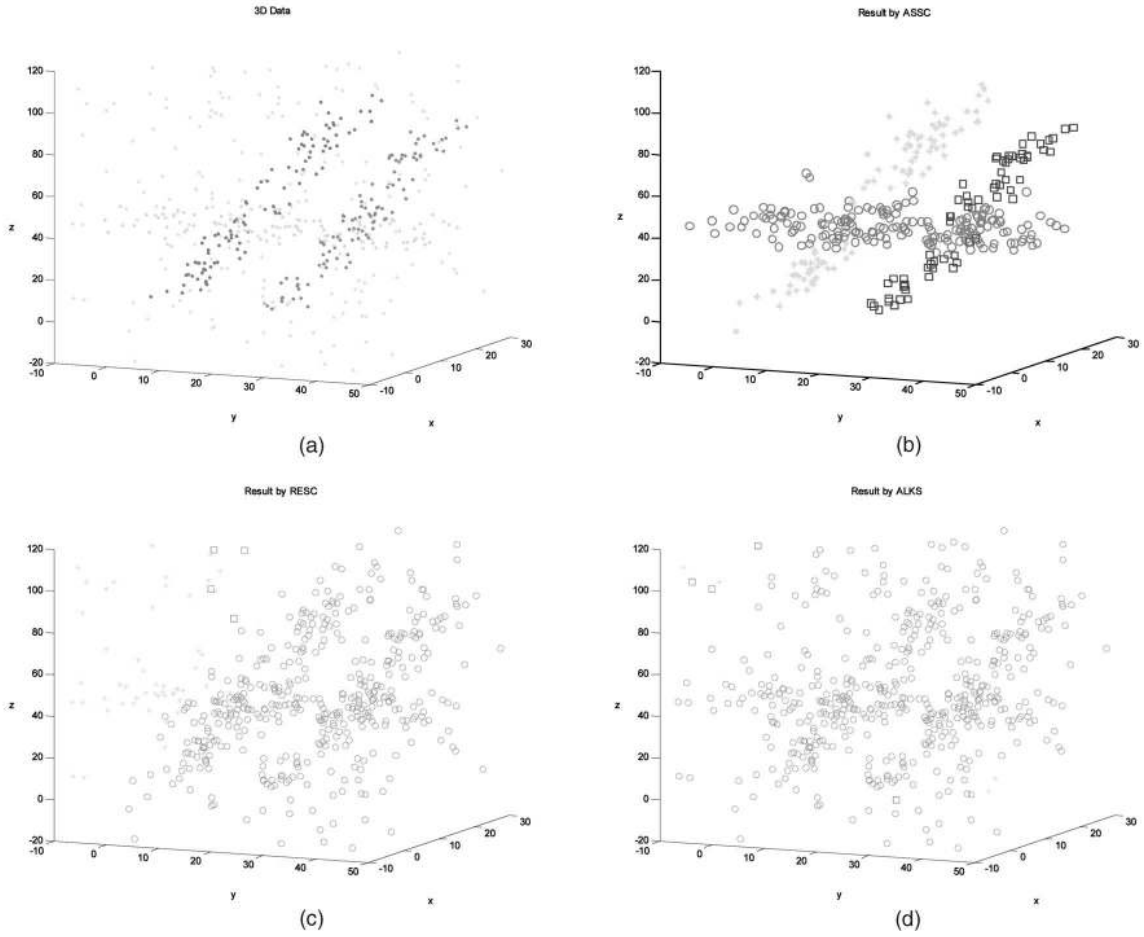

Fig. 6. Second experiment for 3D multiple-structure data: (a) the 3D data; the results by (b) the proposed method, (c) by RESC, (d) and by ALKS.

TABLE 3
Result of the Estimates of the Parameters $(\mathrm{A}, \mathrm{B}, \mathrm{C}; \sigma)$ Provided by Each of the Robust Estimators Applied to the Data in Fig. 6

	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)		

Examples for data with 20 percent and 70 percent outliers are shown in Figs. 7a and 7b to illustrate the distributions of the inliers and outliers. If an estimator is robust enough to outliers, it can resist the influence of both clustered outliers and randomly distributed outliers even when the outliers occupy more than 50 percent of the data. We performed the experiments 20 times, using different random sampling seeds, for each data set involving different percentage of outliers (10 to 90 percent). An averaged result is show in Figs. 7c, 7d, and 7e. From Figs. 7c, 7d, and 7e, we can see that our method obtains the best result. Because the LMedS has only 50 percent breakdown point, it broke down when the outliers approximately occupied more than 50 percent of the data. ALKS broke down when the outliers reached 75 percent. RESC began to break down when the outliers comprised more than 83 percent of the whole data. In contrast, the ASSC estimator is the most robust to outliers. It began to breakdown at 89 percent outliers. In fact, when the inliers are about (or less than) 10 percent of the data, the assumption that inliers should occupy a relative majority of the data is violated. Bridging between the inliers and the clustered outliers tends to yield a higher score. Other robust estimators also suffer from the same problem.

6.4 Influence of the Noise Level of Inliers on the Results of Robust Fitting

Next, we will investigate the influence of the noise level of the inliers. We use the signal shown in Fig. 7b with 70 percent outliers. However, we changed the standard variance of the plane signal from 0.1 to 3, with an increment of 0.1.

Fig. 8 shows that LMedS broke down first. This is because that LMedS cannot resist the influence of outliers when the outliers occupy more than a half of the data points. ALKS, RESC, and ASSC estimators all can tolerate more than 50 percent outliers. However, among these three robust

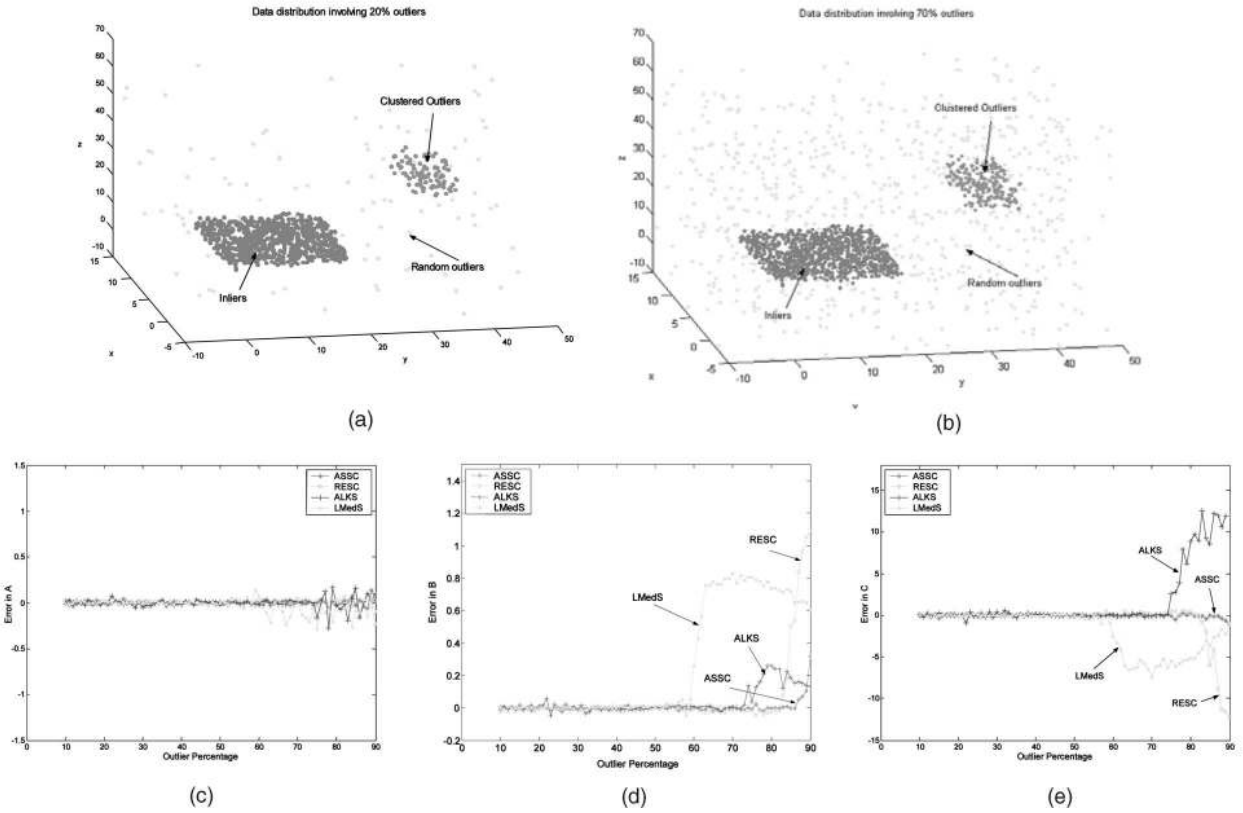

Fig. 7. Error plot of the four methods: (a) example of the data with 20 percent outliers, (b) example of the data with 80 percent outliers, (c) the error in the estimate of parameter A, (d) in parameter B, and (e) in parameter C.

estimators, ALKS broke down first. It began to break down when the noise level of inliers is increased to 1.7. RESC is more robust than ALKS: It began to break down when the noise level of inliers is increased to 2.3. The ASSC estimator shows the best achievement. Even when the noise level is increased to 3.0, the ASSC estimator did not break down yet.

6.5 Influence of the Relative Height of Discontinuous Signals

Discontinuous signals (such as parallel lines/planes, step lines/planes, etc.) often appear in computer vision tasks. Work has been done to investigate the behaviour of robust estimators for discontinuous signals, e.g., [21], [30], [31]. Discontinuous signals are hard to deal with, e.g., most robust estimators break down and yield a "bridge" between the two planes of one step signal. The relative height of the discontinuity is a crucial factor. In this section, we will investigate the influence of the discontinuity on the performance of the four methods.

We generate two discontinuous signals: one containing two parallel planes and one containing one-step planes. The signals have unit variance. Randomly distributed outliers covering the regions of the signals are added to the signals. Among the total 1,000 data points, there are 20 percent pseudo-outliers and 50 percent random outliers. The relative height is increased from 1 to 20. Figs. 9a and 9b show examples of the data distributions of the two signals with relative height 10. The averaged results (over 20 repetitions) obtained by the four robust estimators are shown in Figs. 9c, 9d, 9e, 9f, 9g, and 9h.

From Fig. 9, we can see the tendency to bridge becomes stronger as the step decreases. LMedS shows the worst results among the four robust estimators. For the remaining three estimators (ASSC, ALKS, and RESC) from Figs. 9c, 9d, 9e, 9f, 9g, and 9h, we can see that:

- For the parallel plane signal, the results by ALKS are affected most by the small step. RESC shows better result than ALKS. However, ASSC shows the best result.
- For the step signal, when the step height is small, all of these three estimators are affected. When the step height is increased, all of the three estimators show robustness to the signal. However, ASSC achieves the best results for small step height signals.

In next sections, we will apply the ASSC estimator to "real world" computer vision tasks: range image segmentation and fundamental matrix estimation.

7 ASSC for Range Image Segmentation

Range image segmentation is a complicated task because range images may contain many gross errors (such as sensor noise, etc.), as well as containing multiple structures. Many robust estimators have been employed to segment range images (such as [19], [20], [21], [29], [36], [39]).

7.1 A Hierarchal Algorithm for Range Image Segmentation

Our range image segmentation algorithm is based on the proposed robust ASSC estimator. We employ a hierarchical structure, similar to [36], in our algorithm. Although MDPE in [36] has similar performance to ASSC, MDPE (using a fix

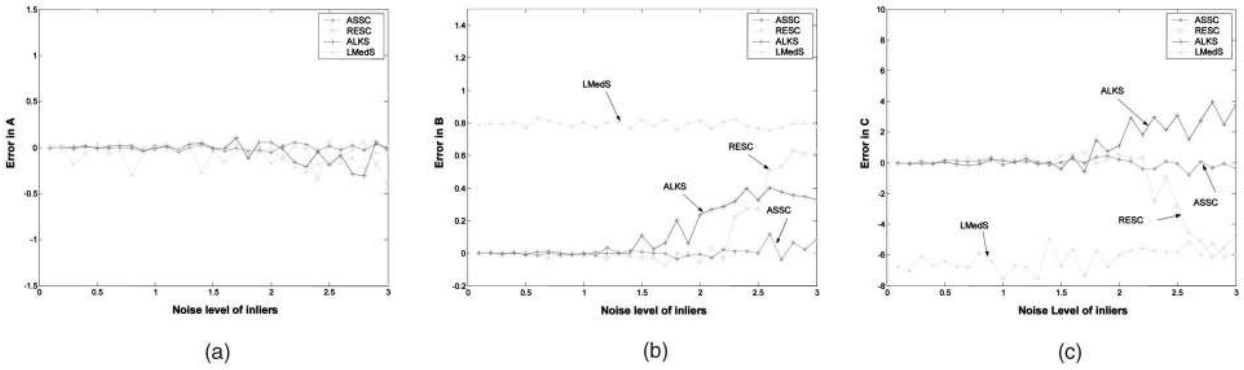

Fig. 8. The influence of the noise level of the inliers on the results of the four methods: plots of the error in the parameters A (a), B (b), and C (c) for different noise levels.

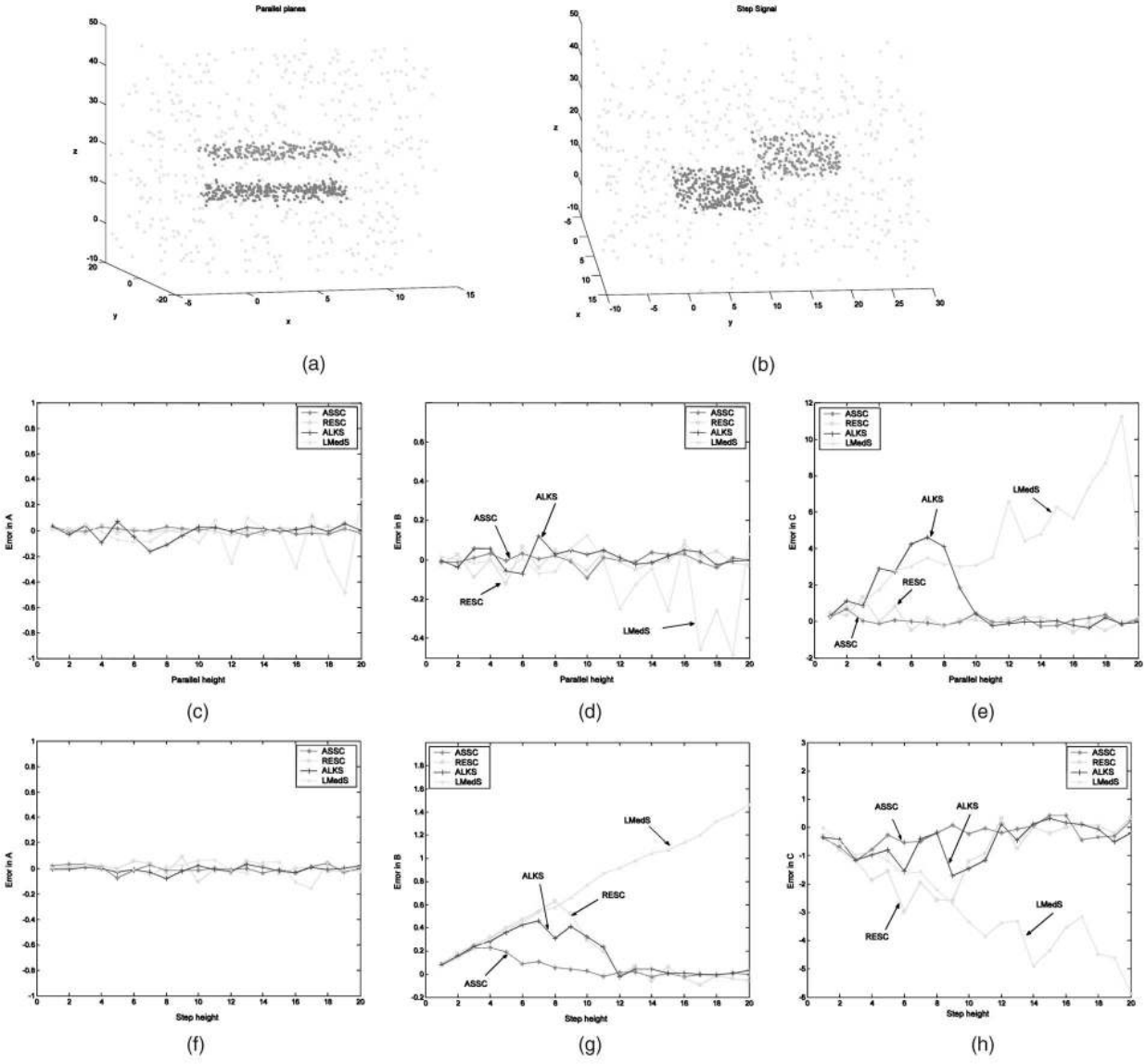

Fig. 9. The influence of the relative height of discontinuous signals on the results of the four methods: (a) two parallel planes; (b) one step signal; (c), (d), and (e) the results for the two parallel planes; (f), (g), and (h) the results for the step signal.

bandwidth technique) only estimates the parameters of the model. An auxiliary scale estimator is needed to provide an estimate of the scale of inliers. ASSC (with a variable bandwidth technique) can estimate the scale of inliers in the process of estimating the parameters of a model. Moreover, ASSC has a more simple objective function.

We apply our algorithm to the ABW range images from the USF database (available at http://marathon.csee.usf.edu/seg-comp/SegComp.html). The range images were generated by using an ABW structured light scanner and all ABW range images have 512 x 512 pixels. Each pixel value corresponds to a depth/range measurement (i.e., in the "z" direction). Thus, the coordinate of the pixel in 3D can be written as (x, y, z), where (x, y) is the image coordinate of the pixel.

Shadow pixels may occur in an ABW range image. These points will be excluded from processing. Thus, all pixels either start with the label "shadow" or they are unlabeled. The basic idea is as follows: From the unlabeled pixels, we find the largest connected component (CCmax). In obtaining the connected components, we do not connect adjacent pixels if there is a "jump edge" between them (defined by a height difference of more than Tjump—a threshold set in advance). We then use ASSC to find a plane within CCmax. The strategy used to speed up the processing is to work on a subsampled version of CCmax, formed by regular subsampling (taking only those points in CCmax that lie on the sublattice spaced by one in r pixels horizontally and one in r pixels vertically), for fitting the plane parameters. This reduces the computation involved in ASSC significantly since one has to deal with significantly less residuals in each application of ASSC.

Note: In the following description, we distinguish carefully between "classify" and "label." Classify is a temporary distinction for the purpose of explanation. Labeling assigns a final label to the pixels, and the set of labels defines the segmentation by groups of pixels with a common label. The steps are:

0. Set r = 8, and set the following thresholds: Tnoise = 5, Tcc = 80, Tinlier = 100, Tjump = 9, Tvalid = 10, Tangle = 45degrees.
1. From the unclassified pixels find the maximum connected component CCmax. If any of the connected components are less than Tnoise in size, we label their data points as "noise." If the number of samples in CCmax is less than Tcc, then go to Step 5. Subsample CCmax to form S. Use ASSC to find the parameters (including scale) of a surface from amongst the data S. Using the parameter found we classify pixels in S into "inlier" or "outlier" and if the number of "inliers" are less than Tinlier then go to Step 5.
2. Using the parameters found in the previous step, classify pixels in CCmax as either "inliers" or "outliers."
3. Classify the "inliers" as either "valid" or "invalid" by the following: When the angle between the normal of the inlier data point and the normal of the estimated plane is less than Tangle, the data point is "valid." If the number of valid points is less than Tvalid, go to Step 5.
4. The "valid" pixels define the region for the fitted plane. Typically, this has holes in it (due to isolated points with large noise). We fill those holes with a general hole filling routine (from Matlab). The pixels belonging to this filled surface are then labeled with a unique label for that surface. Repeat from Step 1 until there are no valid connected components in size (i.e., <Tcc) left to process.
5. Adjust the subsampling ratio: r = r/2 and terminate when r < 1, else go to Step 1.

Finally, we eliminate any remaining isolated outliers and the points labeled as "noise" in Step 1 by assigning them to the majority of their eight connected neighbors.

7.2 Experiments on Range Image Segmentation

Due to the adoption of the robust ASSC estimator, the proposed algorithm is very robust to noise. In this first experiment, we added 26,214 random noise points to the range images taken from the USF ABW range image database (test 7 and train 5). We directly segment the unprocessed raw images. No separate noise filtering is performed.

As shown in Fig. 10, all of the main surfaces were recovered by our method. Only a slight distortion appeared on some boundaries of neighboring surfaces. This is because of the sensor noise and the limited accuracy of the estimated normal at each range point. In fact, the more accurate the range data are, and the more accurate the estimated normals at range points are, the less the distortion is.

We also compare our results with those of several state-of-the-art approaches [16]: University of South Florida (USF), Washington State University (WSU), and University of Edinburgh (UE). (The parameters of these three approaches were trained using the ABW training range images [16].) Figs. 11c, 11d, 11e, and 11f show the results obtained by the four methods. The results by the four methods should be compared with the ground truth (Fig. 11b).

From Fig. 11c, we can see that the USF results contained many noisy points. In Fig. 11d, the WSU segmenter missed one surface; it also over segmented one surface. Some boundaries on the junction of the segmented patch by WSU were relatively seriously distorted. UE shows relatively better results than those of USF and WSU. However, some estimated surfaces are still noisy (see Fig. 11e). Compared with the other three methods, the proposed method achieved the best results. All surfaces are recovered and the segmented surfaces are relatively "clean." The edges of the segmented patches were reasonably good.

Adopting a hierarchical-sampling technique in the proposed method greatly reduces its time cost. The processing time of the method is affected to a relatively large extent by the number of surfaces in the range images. The processing time for a range image including simple objects is faster than that for a range image including complicated objects. Generally speaking, given m = 500, it takes less than one minute (on an AMD800MHz personal computer in C interfaced with the MATLAB language) for segmenting a range image with simple surfaces and about 1-2 minutes for one including complicated surfaces. This includes the time for computing normal information at each range pixel (which takes about 12 seconds).

8 ASSC for Fundamental Matrix Estimation

8.1 Fundamental Matrix Estimation

The fundamental matrix provides a constraint between corresponding points in multiple views. Estimation of the fundamental matrix is important for several problems: matching, recovering of structure, motion segmentation, etc. [34]. Robust estimators such as M-estimators, LMedS, RANSAC, MSAC, and MLESAC have been applied to estimate the fundamental matrix [33].

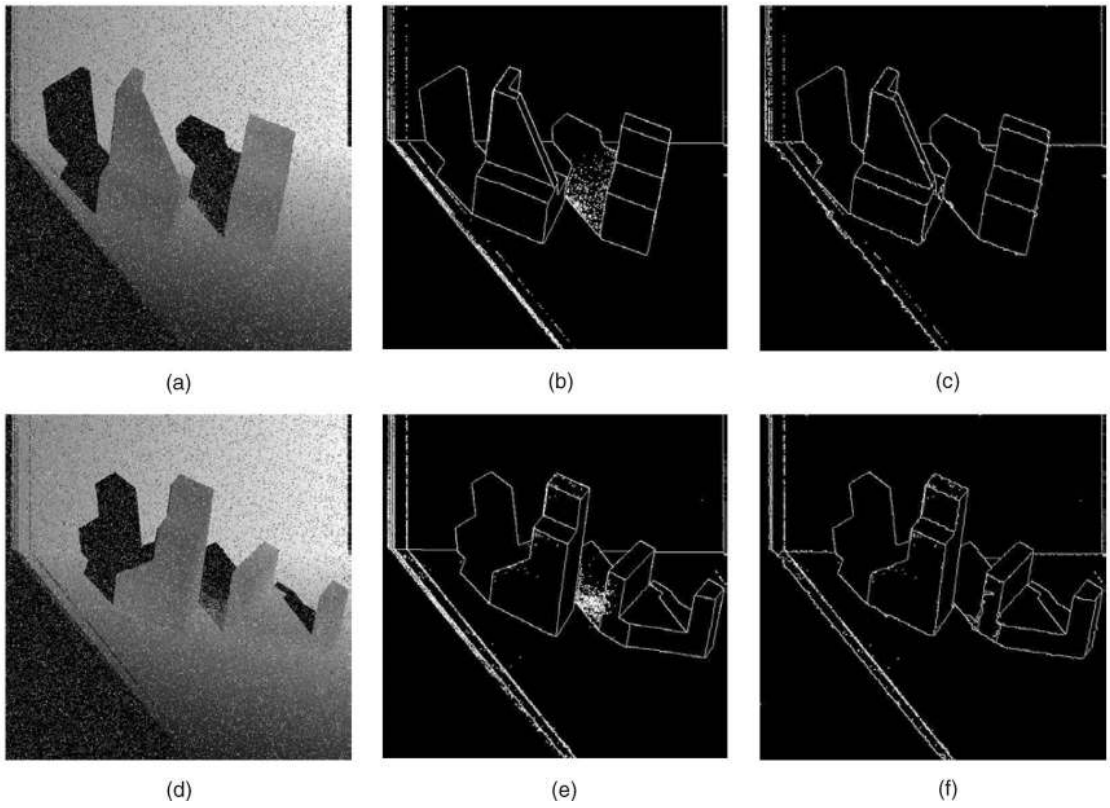

Fig. 10. Segmentation of ABW range images from the USF database. (a), (d) Range image with 26,214 random noise points. (b), (e) The ground truth results for the corresponding range images without adding random noise. (c), (f) Segmentation result by the proposed algorithm.

Let $\{\mathrm{x_i}\}$ and $\{\mathrm{x_i'}\}$ (for $\mathrm{i} = 1, \ldots, \mathrm{n}$) be a set of matched homogeneous image points viewed in image 1 and image 2, respectively. We have the following constraints for the fundamental matrix F:

$$x_i^{'T} F x_i = 0 \; and \det[F] = 0. \tag{24}$$

We employ the 7-point algorithm ([33, p. 7]), to solve for candidate fits using Simpson distance. For the $i\mathrm{th}$ correspondence, residual r_i using Simpson distance is:

$$r_i = \frac{k_i}{\left(k_x^2 + k_y^2 + k_{x'}^2 + k_{y'}^2\right)^{1/2}}, \tag{25}$$

where

$$\begin{aligned} k_i =& f_1 x_i' x_i + f_2 x_i' y_i + f_3 x_i' \varsigma + f_4 y_i' x_i \\ &+ f_5 y_i' y_i + f_6 y_i' \varsigma + f_7 x_i \varsigma + f_8 y_i \varsigma + f_9 \varsigma^2. \end{aligned}$$

8.2 The Experiments on Fundamental Matrix Estimation

First, we generated 300 matches including 120 point pairs of inliers with unit Gaussian variance and 160 point pairs of random outliers. In practice, the scale of the inliers is not available. Thus, the median scale estimator, as recommended in [33], is used for RANSAC and MSAC to yield an initial scale estimate. The number of random samples is set to 10,000. The experiment was repeated 30 times and the averaged values are shown in Table 4. From Table 4, we can see that our method yields the best result.

Next, we draw the *error plot* of the four methods. Among the total 300 correspondences, the percentage of outliers is increased from 5 to 70 percent in increments of 5 percent. The experiments were repeated 100 times for each percentage of outliers. If a method is robust enough, it should resist the influence of outliers and the correctly identified percentages of inliers should be around 95 percent (T is set 1.96 in (1)) and the standard variance of inliers should be near to 1.0 regardless of the percentages of outliers actually in the data. We set the number of random samples, m, to be high enough to ensure a high probability of success.

From Fig. 12, we can see that MSAC, RANSAC, and LMedS all break down when the data involve more than 50 percent outliers. The standard variance of inliers by ASSC is the smallest when the percentage of outliers is higher than 50 percent. Note: ASSC succeeds to find the inliers and outliers even when the outliers occupied 70 percent of the whole data. Finally, we apply the proposed method on real image frames: two frames of the Corridor sequence (bt.003 and bt.006), which can be obtained from http://www.robots.ox.ac.uk/~vgg/data/ (Figs. 13a and 13b). Fig. 13c shows the matches involving 500 point pairs in total. The inliers (201 correspondences) obtained by the proposed method are shown in Fig. 13d. The epipolar lines (we draw 30 of the epipolar lines) and the epipole found using the estimated fundamental matrix by ASSC are shown in Figs. 13e and 13f. We can see that the proposed method achieves a good result. Because the camera matrices of the two frames are available, we can obtain the ground truth fundamental matrix and, thus, evaluate the errors. From Table 5, we can see that ASSC performs the best among the four methods.

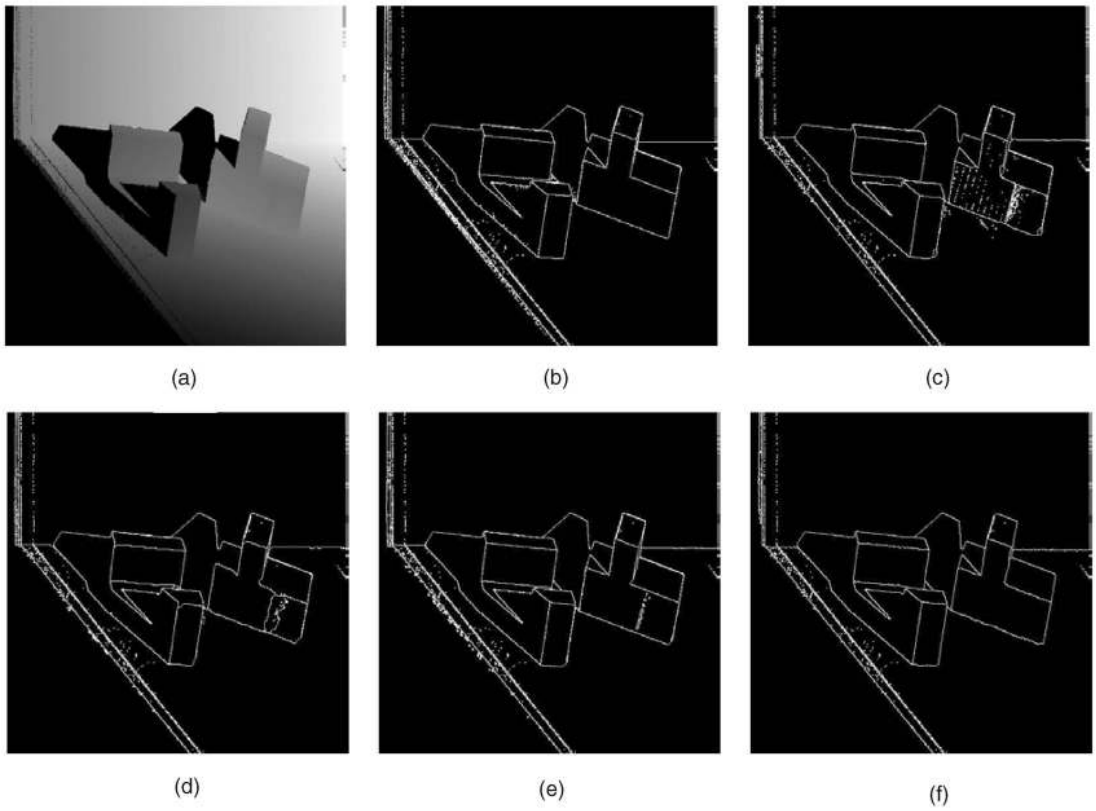

Fig. 11. Comparison of the segmentation results for ABW range image (test 3). (a) Range image. (b) The result of ground truth. (c) The result by the USF. (d) The result by the WSU. (e) The result by the UE. (f) The result by the proposed method.

TABLE 4
An Experimental Comparison for Data with 60 Percent Outliers

	% of inliers correctly classified	% of outliers correctly classified	Standard variance of inliers
Ground Truth	100.00	100.00	1.0
ASSC	96.08	98.43	1.4279
MSAC	100.00	60.48	75.2288
RANSAC	100.00	0.56	176.0950
LMedS	100.00	61.91	65.1563

9 Conclusion

We have shown that scale estimation for data, involving multiple structures and high percentages of outliers, is as yet a relatively unsolved problem. As a partial solution, we introduced a robust two-step scale estimator (TSSE) and we presented experiments showing its advantages over other existing robust scale estimators. TSSE can be used to give an initial scale estimate for robust estimators such as M-estimators, etc. TSSE can also be used to provide an auxiliary estimate of scale (after the parameters of a model have been found) as a component of almost any robust fitting method.

We also proposed a very robust Adaptive Scale Sample Consensus (ASSC) estimator which has an objective function that takes account of both the number of inliers and the corresponding scale estimate for those inliers. ASSC is very robust to multiple-structural data containing high percentages of outliers (more than 80 percent outliers). The ASSC estimator was compared to several popular robust estimators and generally achieves better results.

Finally, we applied ASSC to range image segmentation and to fundamental matrix estimation. However, the applications of ASSC are not limited to these two fields. The computational cost of the proposed ASSC method is moderately low.

Although we have compared against several of the "natural competitors" from the computer vision literature, it is difficult to be comprehensive. For example, in [26], the authors also proposed a method which can simultaneously estimate the model parameters and the scale of the inliers. In essence, the method tries to find the fit that produces residuals that are the most Gaussian distributed (or which have a subset that is most Gaussian distributed), and all data points are considered. In contrast, only the data points, within the band obtained by the mean shift and mean shift valley, are considered in our objective function. Also, we do not assume that the residuals for the best fit will be the best match to a

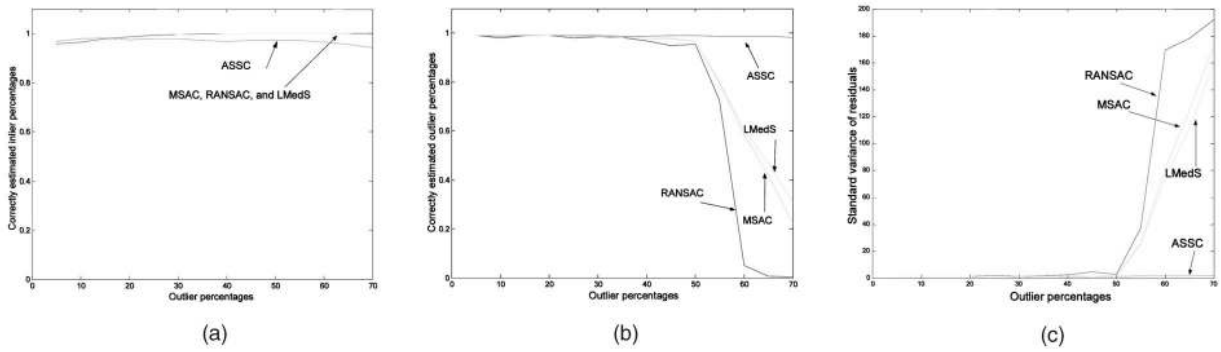

Fig. 12. A comparison of the correctly identified percentage of inliers (a), of outliers (b), and of the standard variance of residuals of inliers (c).

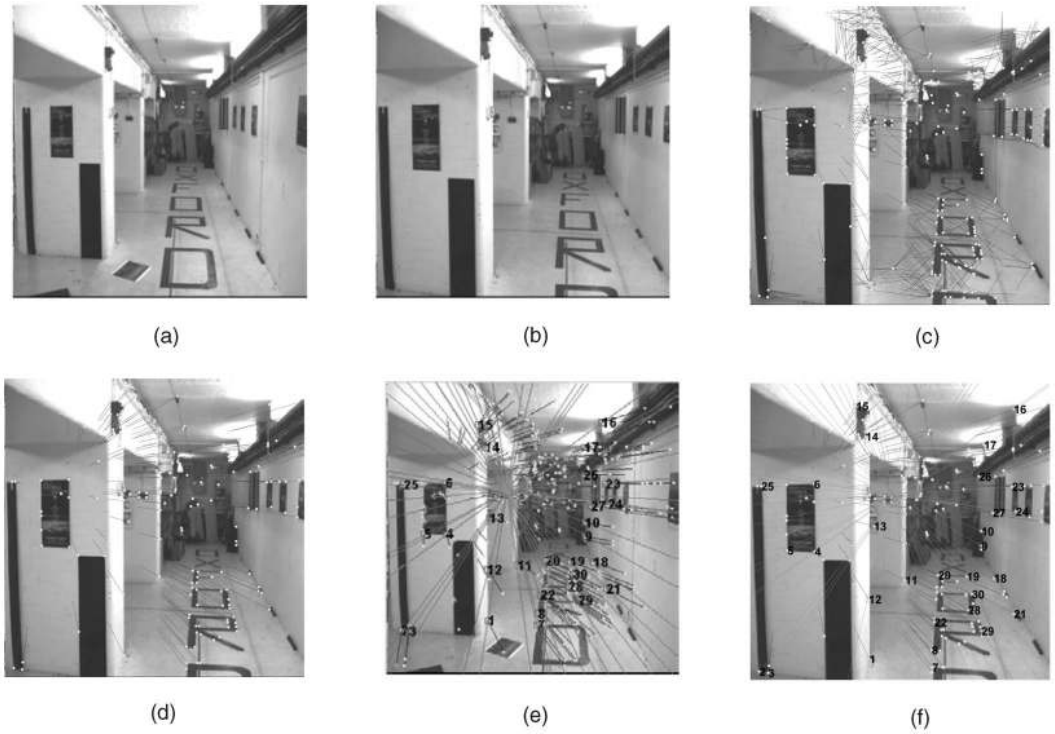

Fig. 13. (a) and (b) image pair, (c) matches, (d) inliers by ASSC; (e) and (f) epipolar geometry.

TABLE 5
Experimental Results on Two Frames of the Corridor Sequence

	Number of inliers (i.e., matched correspondences)	Mean error of inliers	Standard variance of inliers
ASSC	201	0.0274	0.6103
MSAC	391	-1.3841	10.0091
RANSAC	392	-1.0887	10.1394
LMedS	372	-1.4921	9.5463

Gaussian distribution. In the latter stage of the preparation of this paper, we become aware that Gotardo et al. [15] proposed an improved robust estimator based on RANSAC and MLESAC, and applied it to range image segmentation. However, like RANSAC, this estimator also requires the user to set the scale-related tolerance. In contrast, the proposed ASSC method does not require any priori information about the scale. Although [4] also employs kernel density estimation technique, it uses the projection pursuit paradigm. Thus, for higher dimensions, the computational complexity is greatly increased. ASSC considers the density distribution of the mode in 1D residual space, by which the dimension of the space is reduced.

A more comprehensive set of comparisons would be a useful piece of future work.

ACKNOWLEDGMENTS

This work is supported by the Australia Research Council (ARC), grant A10017082. The authors would like to thank Professor Andrew Zisserman, Dr. Hongdong Li, Kristy Sim, and Haifeng Chen for their kind help in relation to aspects

of fundamental matrix estimation. They also thank the anonymous reviewers for their valuable comments including drawing our attention to [26].

REFERENCES

[1] A. Bab-Hadiashar and D. Suter, "Robust Optic Flow Computation," *Int'l J. Computer Vision,* vol. 29, no. 1, pp. 59-77, 1998.

[2] A. Bab-Hadiashar and D. Suter, "Robust Segmentation of Visual Data Using Ranked Unbiased Scale Estimate," *ROBOTICA, Int'l J. Information, Education and Research in Robotics and Artificial Intelligence,* vol. 17, pp. 649-660, 1999.

[3] M.J. Black and A.D. Jepson, "EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation," *Int'l J. Computer Vision,* vol. 26, pp. 63-84, 1998.

[4] H. Chen and P. Meer, "Robust Computer Vision through Kernel Density Estimation," *Proc. European Conf. Computer Vision,* pp. 236-250, 2002.

[5] H. Chen and P. Meer, "Robust Regression with Projection Based M-Estimators," *Proc Ninth Int'l Conf. Computer Vision,* 2003.

[6] H. Chen, P. Meer, and D.E. Tyler, "Robust Regression for Data with Multiple Structures," *Proc. 2001 IEEE Conf. Computer Vision and Pattern Recognition (CVPR),* 2001.

[7] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 8, pp. 790-799, 1995.

[8] D. Comaniciu and P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation," *Proc. 1997 IEEE Conf. Computer Vision and Pattern Recognition,* pp. 750-755, 1997.

[9] D. Comaniciu and P. Meer, "Mean Shift Analysis and Applications," *Proc. Seventh Int'l Conf. Computer Vision,* pp. 1197-1203, 1999.

[10] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach towards Feature Space A Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 5, pp. 603-619, May 2002.

[11] D. Comaniciu, V. Ramesh, and A.D. Bue, "Multivariate Saddle Point Detection for Statistical Clustering," *Proc. Europe Conf. Computer Vision (ECCV),* 2002.

[12] D. Comaniciu, V. Ramesh, and P. Meer, "The Variable Bandwidth Mean Shift and Data-Driven Scale Selection," *Proc. Eighth Int'l Conf. Computer Vision,* 2001.

[13] M.A. Fischler and R.C. Rolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Comm. ACM,* vol. 24, no. 6, pp. 381-395, 1981.

[14] K. Fukunaga and L.D. Hostetler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition," *IEEE Trans. Information Theory,* vol. 21, pp. 32-40, 1975.

[15] P. Gotardo, O. Bellon, and L. Silva, "Range Image Segmentation by Surface Extraction Using an Improved Robust Estimator," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2003.

[16] A. Hoover, G. Jean-Baptiste, and X. Jiang, "An Experimental Comparison of Range Image Segmentation Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 7, pp. 673-689, July 1996.

[17] P.V.C. Hough, "Methods and Means for Recognising Complex Patterns," US Patent 3 069 654, 1962.

[18] P.J. Huber, *Robust Statistics.* Wiley, 1981.

[19] K. Koster and M. Spann, "MIR: An Approach to Robust Clustering—Application to Range Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 5, pp. 430-444, May 2000.

[20] K.-M. Lee, P. Meer, and R.-H. Park, "Robust Adaptive Segmentation of Range Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 2, pp. 200-205, Feb. 1998.

[21] J.V. Miller and C.V. Stewart, "MUSE: Robust Surface Fitting Using Unbiased Scale Estimates," *Proc. Conf. Computer Vision and Pattern Recognition,* 1996.

[22] E.P. Ong and M. Spann, "Robust Optical Flow Computation Based on Least-Median-of-Squares Regression," *Int'l J. Computer Vision,* vol. 31, no. 1, pp. 51-82, 1999.

[23] P.J. Rousseeuw, "Least Median of Squares Regression," *J. Am. Statistics Assoc.,* vol. 79, pp. 871-880, 1984.

[24] P.J. Rousseeuw and C. Croux, "Alternatives to the Median Absolute Derivation," *J. Am. Statistical Associaion,* vol. 88, no. 424, pp. 1273-1283, 1993.

[25] P.J. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection.* John Wiley & Sons, 1987.

[26] D.W. Scott, "Parametric Statistical Modeling by Minimum Integrated Square Error," *Technometrics,* vol. 43, no. 3, pp. 274-285, 2001.

[27] A.F. Siegel, "Robust Regression Using Repeated Medians," *Biometrika,* vol. 69, pp. 242-244, 1982.

[28] B.W. Silverman, *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 1986.

[29] C.V. Stewart, "MINPRAN: A New Robust Estimator for Computer Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 10, pp. 925-938, Oct. 1995.

[30] C.V. Stewart, "Bias in Robust Estimation Caused by Discontinuities and Multiple Structures," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 8, pp. 818-833, Aug. 1997.

[31] C.V. Stewart, "Robust Parameter Estimation in Computer Vision," *SIAM Rev.,* vol. 41, no. 3, pp. 513-537, 1999.

[32] G.R. Terrell and D.W. Scott, "Oversmoothed Nonparametric Density Estimates," *J. Am. Statistical Association,* vol. 80, pp. 209-214, 1985.

[33] P. Torr and D. Murray, "The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix," *Int'l J. Computer Vision,* vol. 24, pp. 271-300, 1997.

[34] P. Torr and A. Zisserman, "MLESAC: A New Robust Estimator With Application to Estimating Image Geometry," *Computer Vision and Image Understanding,* vol. 78, no. 1, pp. 138-156, 2000.

[35] M.P. Wand and M. Jones, *Kernel Smoothing.* Chapman & Hall, 1995.

[36] H. Wang and D. Suter, "MDPE: A Very Robust Estimator for Model Fitting and Range Image Segmentation," *Int'l J. Computer Vision,* to appear.

[37] H. Wang and D. Suter, "False-Peaks-Avoiding Mean Shift Method for Unsupervised Peak-Valley Sliding Image Segmentation," *Digital Image Computing Techniques and Applications,* pp. 581-590, 2003.

[38] H. Wang and D. Suter, "Variable Bandwidth QMDPE and Its Application in Robust Optic Flow Estimation," *Proc. Int'l Conf. Computer Vision,* pp. 178-183, 2003.

[39] X. Yu, T.D. Bui, and A. Krzyzak, "Robust Estimation for Range Image Segmentation and Reconstruction," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 16, no. 5, pp. 530-538, May 1994.

[40] Z. Zhang et al., "A Robust Technique for Matching Two Uncalibrated Image through the Recovery of the Unknown Epipolar Geometry," *Artificial Intelligence,* vol. 78, pp. 87-119, 1995.

Hanzi Wang received the BSc degree in physics and the MSc degree in optics from Sichuan University, China, in 1996 and 1999, respectively. He is currently a PhD candidate in the Department of Electrical and Computer Systems Engineering at Monash University, Australia. His current research interest are mainly concentrated on computer vision and pattern recognition including robust statistics, model fitting, optical flow calculation, image segmentation, fundamental matrix estimation, and related fields. He has published more than 10 papers in major international journals and conferences.

David Suter received the BSc degree in applied mathematics and physics from The Flinders University of South Australia (1977) and the PhD degree in computer vision from La Trobe University (1991). His main research interest is motion estimation from images and visual reconstruction. He is an associate professor in the Department of Electrical and Computer Systems Engineering at Monash University, Australia. He served as general cochair of the 2002 Asian Conference in Computer Vision and has been cochair of the Statistical Methods in Computer Vision Workshops (2002 Copenhagen and 2004 Prague). He currently serves on the editorial board of two international journals: *The International Journal of Computer Vision* and *The International Journal of Image and Graphics*. He is a senior member of the IEEE and vice president of the Australian Pattern Recognition Society.