Department of Electrical
and
Computer Systems Engineering

Technical Report
MECSE-2-2003

Robust Scale Estimation from True Parameters of Model

H. Wang and D. Suter

MONASH
UNIVERSITY

Robust Scale Estimation from True Parameters of Model

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

Abstract

In computer vision tasks, it frequently happens that gross noise and pseudo outliers occupy the absolute majority of the data. During the past several decades, a lot of robust estimators were developed to find parameters of a model from heavily contaminated data. However, correctly estimating the parameters of a model is not enough to differentiate inliers from outliers. Robust scale estimation is often needed as the postprocessing of most robust estimators followed by a weighted least squares method on the inliers. This paper shows that the scale estimation for most robust estimators is a very weak field and more work is needed. A more robust two-step scale estimator is presented and comparative experiments show its advantages over other available scale estimators.

1. Introduction

One major task of pattern recognition, machine learning, and related areas: is to fit a model to noisy data (with outliers). It is common to employ "regression analysis" to undertake such tasks. The most common form of regression analysis is the least squares (LS) method, which can achieve optimum results under Gaussian distributed noise. But this method is extremely sensitive to outliers (gross errors or samples belonging to another structure and distribution). The breakdown point of an estimator may be roughly defined as the smallest percentage of outlier contamination that can cause the estimator to produce arbitrarily large values. 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%.

Since data contamination is usually unavoidable (due to faulty feature extraction, sensor noise and failure, segmentation errors, etc.), there has recently been a general recognition that algorithms should be robust [1]. Robust regression methods are a class of techniques that can tolerate gross errors (outliers). Some robust methods also have a high breakdown point.

Most past work mainly aimed at presenting robust estimators with high breakdown point [2-6], i.e. the estimator can correctly find the parameters of a model from the data which are heavily contaminated. However, correctly estimating the parameters of a model is not enough to differentiate inliers from outliers. One frequently needs an initial or auxiliary estimate of scale, for example, Hough transform needs an auxiliary estimate of scale after finding the parameters of the model, so that the inliers can be differentiated from outliers. Robust scale estimation is often carried out as a postprocessing of most robust estimators to tell the inliers from outliers. A weighted least squares method is then employed on the inliers. Whether or not the inliers can be successfully differentiated from the outliers depends on (1) whether the parameters of a model are correctly found; (2) whether the scale of inliers is correctly estimated.

In this paper, we assume we have found the true parameters of the model to fit. We investigate the behavior of several robust scale estimators that are widely used in computer vision community and show the internal problems of these scale estimation techniques. More work in this field is needed.

This paper is organized as follows: in section 2, we review previous robust scale techniques, and proposed a novel robust scale estimator. Comparative experiments are contained in section 3. We conclude in section 4.

2. 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 [7]:

$$1.4826\,\mathrm{med}_i x_i \qquad (1)$$

The inliers are the points that satisfy the following condition:

$$x_j / 1.4826\,\mathrm{med}_i x_i < T \qquad (2)$$

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

MAD is often used to estimate the scale of inliers. It has a simple explicit formula and is computationally efficient [8]:

$$\mathrm{MAD} = 1.4826\,\mathrm{med}_i\{|x_i - \mathrm{med}_j x_j|\} \qquad (3)$$

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

$$\frac{\left|x_i - med_j x_j\right|}{MAD_n} \qquad (4)$$

When equation (4) for a point x_i dranw from a sample $\{x_j\}$ exceeds a threshold, say 2.5, an outlier is recognized.

The median and the MAD are often used as initial values for more robust estimators. The two estimators can also serve as ancillary scale estimators for other more robust estimators.

Because the median and the MAD have 50% breakdown points, it means they will break down when data including more than 50% outliers. Another problem we found is that both the median and the MAD scale estimators are biased even when data contains less than 50% outliers (see section 3).

2.2. Adaptive Least K-th Squares (ALKS) Estimator

The authors of ALKS [5] consider robust 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 scale estimate, assuming inliers have a Gaussian distribution, is given:

$$\hat{s}_k = \frac{\hat{d}_k}{\Phi^{-1}[(1+k/n)/2]} \qquad (5)$$

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 correspond to the minimum of the variance of the normalized error ε_k^2:

$$\varepsilon_k^2 = \frac{1}{k-p}\sum_{i=1}^{k}\left(\frac{r_{i,k}}{\hat{s}_k}\right)^2 = \frac{\hat{\sigma}_k^2}{\hat{s}_k^2} \qquad (6)$$

They assumed that when k is increased so that 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 in ALKS is its lack of stability under a small percentage of outliers (which will be illustrated in section 3).

2.3. Modified Selective Statistical Estimator (MSSE)

Bab-Hadiashar and Suter [9] have used least k-th order (rather than median) methods and a heuristic way of estimating scale to perform range segmentation. After finding a fit, they tried to recognize the first outlier, by detecting the k-th residual jumps, which can indicate the unbiased scale estimate using the first k-th residuals in an ascending order:

$$\hat{\sigma}_k^2 = \frac{\sum_{i=1}^{k} r_i^2}{k-p} \qquad (7)$$

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. Thus, the scale can be estimated by checking the validity of the following inequality:

$$\frac{\sigma_{k+1}^2}{\sigma_k^2} > 1 + \frac{T^2-1}{k-p+1} \qquad (8)$$

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

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 [3]. 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. After finding a fit, they estimate the scale of the fit by directly calculating the follows:

$$\sigma = \alpha\left(\frac{1}{\sum_{i=1}^{v} h_i^c - 1}\sum_{i=1}^{v}(ih_i^c\delta - \overline{h}^c)^2\right)^{1/2} \qquad (9)$$

where $\overline{h}^c$ 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 found the estimated scale is still overestimated for the reason that, instead of summing up squared differences between all individual residuals and the mean residual in the compressed histogram, equation (8) sums up the squared differences between

residuals in each bin of compressed histogram and the mean residual in the compressed histogram.
We revise it as follows:

$$\sigma = \alpha\left(\frac{1}{\sum_{i=1}^{v} h_i^c - 1}\sum_{i=1}^{n_c}(r_i - \overline{h}^c)^2\right)^{1/2} \quad (10)$$

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

2.5. Two-step scale estimator (TSSE)

We have observed that the median and the MAD is biased when the data include multiple modes. This is because both estimators are provided assuming the whole data have a Gaussian distribution. We base our method on the assumption that the inliers occupy relative majority, and are Gaussian distributed, but the whole data can include multiple-structural distribution. Thus, we propose a robust two-step method to estimate the scale of the inliers.

(1) Because the inliers have Gaussian distribution, we use mean shift, with initial center zero, to find the local peak, and then we use the mean shift valley [10] to find the valley near to the peak. All these are performed in signed ascending ordered residual space. Thus we obtain the half-bandwidth of window centered at the local peak. Other modes other than the inliers will be disregarded outside the obtained window.
(2) Then we estimate the scale of the fit by the median scale estimator on the points within the obtained window centered at the local peak.

In next section, we will compare the achievements of our method and other methods. The experiments will show the advantages of the proposed method over other methods.

3. Experiments

In the following experiments, 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).

3.1. Normal distribution

In this subsection, we generate a simple line signal:
One line: x:(0-55), y=30, γ_1=10000, σ_1=3; α=0;
We use the median (1), the MAD (2), the ALKS (3) , the MSSE (4), the revised RESC (5), and the TSSE (6) to estimate the scale of the line signal.
As results, we obtained the median (3.0258); the MAD (3.0237); the ALKS (2.0061); the MSSE (2.8036); the revised RESC (2.8696); and the TSSE (3.0258). Among these six comparative methods, the median, the MAD, and the TSSE gave the most accurate results. The 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 (6) wrongly estimates the optimal k. It used only about 15% data as inliers. The MSSE used 98% data points as inliers, which is reasonable good.

3.2 Two-mode distribution with random noise

In this subsection, we will use relatively complicated data. We generated a step signal so that the data include two modes.
A step signal: line1: x:(0-55), y=30, γ_1=3000, σ_1=3; line2: x:(55-100), y=40, γ_2=2000, σ_2=3; α=0.
We obtained: the median (6.0432); the MAD (8.6817); the ALKS (3.1823); the MSSE (2.7792); the revised RESC (2.8251); and the TSSE (2.8765). Among these six comparative methods, the median and the MAD gave the worst results. This is because the median and the MAD scale estimators assume the residuals of the whole data are at Gaussian distribution. The other four scale estimators yield good results.

3.3 Two-mode distribution with more outliers

In this subsection, we still 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=750; γ_2=500; α=3250.
After applying the six methods to estimate the scale of the signal, we obtained: the median (37.7385); the MAD (30.6652); the ALKS (23.6490); the MSSE (31.8886); the revised RESC (25.4960); and the TSSE (4.7843).
From the obtained results, we can see that only our proposed method gave a good result (reasonably good), while all other five methods failed to estimate the scale of the inliers when the data involve a high percentage of outliers.

3.4 Breakdown plot

3.4.1 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=50, and α=400, i.e. the data include 90% outliers. The results are repeated 20 times.
Figure 1 shows that TSSE yielded the best results among the six comparative methods. The revised RESC begin to breakdown when the outliers have more than about 70%. The 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. The ALKS gave less accurate results than TSSE.

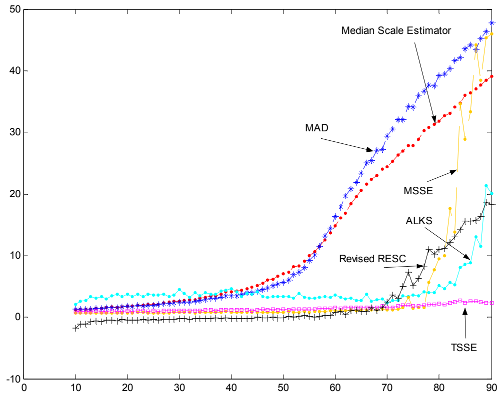

Fig. 1. Breakdown plot of six methods in estimating the scale of a roof signal.

3.4.2 A step signal

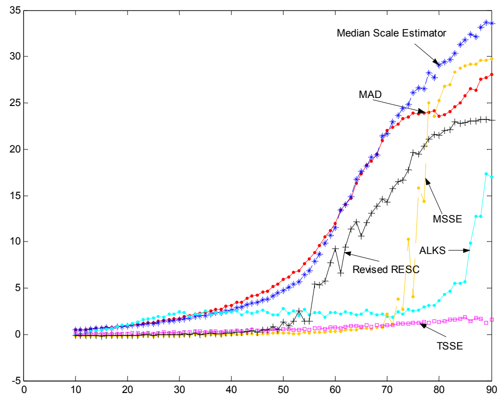

Fig. 2. Breakdown plot of six methods in estimating the scale of a step signal.

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

One-step signal: x:(0-55), y=30, γ_1, σ=3; x:(55-100), y=40, γ_2=100; σ=3.

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=100, and α=800, i.e. the data include 90% outliers.

From figure 1, we can see TSSE gave the most accurate estimation of the scale of the signal. The revised RESC begin to breakdown when the outliers have more than about 50%. The MSSE gave reasonable results when the percentage of outliers is less than 70%, but it broke down when the data include more outliers. The median and the MAD scale estimators are more biased with the increase in the percentage of outliers. The ALKS has less accurate results than TSSE and, when the percentage of outliers is less than 50%, it has less accurate results than the revised RESC, and the MSSE.

Compared with figure 1, we can see increasing the true scale of inliers will lead to less accurate results of the revised RESC, MSSE, and ALKS, but it has less influence on the results of the proposed TSSE. Even when the data include 90% outliers, the TSSE recovered the scale of inliers: 4.32 for the roof and 4.57 for the step signal, which is reasonably good.

3.4.3 Breakdown plot for robust scale estimator

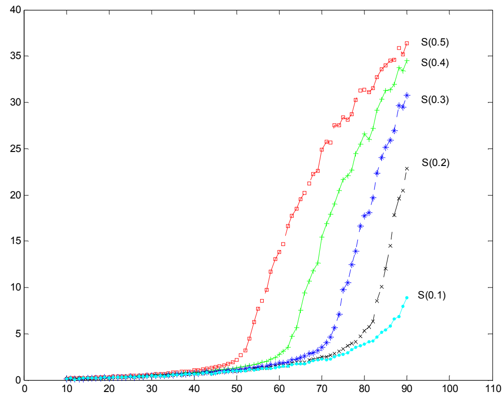

Fig.3 Breakdown plot of different robust scale estimator

If the data have a Gaussian distribution, the median scale estimator (1) is only one case of the robust scale estimator (5). Although theory has proved that the least kth squares method has only min(k/n, 1-k/n) breakdown point, it would be interesting to investigate the achievements of the robust scale estimator after the correct parameters of a model have been found. We let

$$S(q)=\frac{\hat{d}_{qn}}{\Phi^{-1}[(1+q)/2]} \qquad (11)$$

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. Finally, γ_1=50, and α=400, i.e. the data include 90% outliers.

As figure 3 shows, after finding the correct 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 provide a useful cue for robust estimators, which use the median scale method to recovery the scale of inliers.

4. Conclusions

In this paper, we investigate the achievements of several robust scale estimators. We find that they are vulnerable when the data included a high percentage of outliers and the scale of the true fit is large. This provides an important warning to the computer vision community: to carefully choose a proper scale estimator is necessary.

We also propose a promising novel scale estimator (TSSE). The experiments show the advantages of the proposed method over other existing methods, especially, when the data involve a high percentage of outliers and the noise level of inliers is large. The TSSE can also be used as an auxiliary estimate of scale by other robust fitting methods such as Hough Transform [11], MDPE [12], etc.

References:

1. Haralick, R.M., *Computer vision theory: The lack thereof.* Comput. Vision Graphics Image Processing, 1986. **36**: p. 372-386.
2. Stewart, C.V., *MINPRAN: A New Robust Estimator for Computer Vision.* IEEE Trans. Pattern Analysis and Machine Intelligence, 1995. **17**(10): p. 925-938.
3. 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): p. 530-538.
4. Rousseeuw, P.J., *Least Median of Squares Regression.* J. Amer. Stat. Assoc, 1984. **79**(871-880).
5. 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): p. 200-205.
6. 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**.
7. Rousseeuw, P.J. and A. Leroy, *Robust Regression and outlier detection.* John Wiley & Sons, New York., 1987.
8. Rousseeuw, P.J. and C. Croux, *Alternatives to the Median Absolute Deriation.* Journal of the American Statistical Associaion, 1993. **88**(424): p. 1273-1283.
9. 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**: p. 649-660.
10. Hanzi Wang and D. Suter. *False-Peaks-Avoiding Mean Shift Method for Unsupervised Peak-Valley Sliding Image Segmentation*. in *Submitted to CVPR 2003*.
11. Hough, P.V.C., *Methods and means for recognising complex patterns*, in *U.S. Patent 3 069 654*. 1962.
12. Hanzi Wang and D. Suter, *MDPE: A Very Robust Estimator for Model Fitting and Range Image Segmentation.* Submitted to IJCV 2002.