ADAPTIVE SCALE BASED ENTROPY-LIKE ESTIMATOR FOR ROBUST FITTING

*Jinlong Cai and Hanzi Wang**

Center for Pattern Analysis and Machine Intelligence, Xiamen University, China
Email: cai.jinlong00@gmail.com, hanzi.wang@xmu.edu.cn

ABSTRACT

In this paper, we propose a novel robust estimator, called ASEE (Adaptive Scale based Entropy-like Estimator) which minimizes the entropy of inliers. This estimator is based on IKOSE (Iterative Kth Ordered Scale Estimator) and LEL (Least Entropy-Like Estimator). Unlike LEL, ASEE only considers inliers' entropy while excluding outliers, which makes it very robust in parametric model estimation. Compared with other robust estimators, ASEE is simple and computationally efficient. From the experiments on both synthetic and real-image data, ASEE is more robust than several state-of-the-art robust estimators, especially in handling extreme outliers.

***Index Terms*—** robust statistics, model fitting, scale estimation, entropy

1. INTRODUCTION

Robust parametric model estimation techniques in computer vision underlie numerous applications such as motion segmentation [1], range image segmentation [2], homography estimation [3], fundamental matrix estimation [4], etc. The key point of those robust methods is in that how to tolerate the influence of noise and outliers, including pseudo-outliers (i.e., structured outliers) and gross outliers.

In order to be robust to noise and outliers, many robust methods have been proposed. The M-estimatros [5] and the RANdom SAmple Consensus (RANSAC) estimator [6] are the two widely used estimators. However, the M-estimators can not tolerate more than 50% outliers; RANSAC can deal with the data with more than 50% outliers, but its robustness depends on an appropriate choice of an error tolerance threshold. M-estimator SAmple Consensus (MSAC) [7] improves the performance of RANSAC by changing its cost function but it also needs the user to specify the error tolerance. Adaptive Least Kth Order Squares (ALKS) [8], Residual Sample Consensus (RESC) [9], REsidual CONsensus (RECON) [4], Adaptive Scale Sample Consensus (ASSC) [10], and Adaptive Scale Kernel Consensus (ASKC) [11], etc., all claim to be able to deal with more than 50% outliers. However, ALKS can not handle extreme outliers; RESC needs the user to tune many parameters in compressing a histogram; RECON is very efficient but not effective in handling structured outliers; ASSC claims to be able to handle extreme outliers, but it is less effective in estimating model parameters because it assigns equal weights to all inliers. Both ASSC and ASKC employ the nonparametric kernel density estimation techniques to estimate the inlier scale of data (i.e., the TSSE scale estimator), which is computationally expensive in finding local peak and local valley in residual space.

Relation to prior work: In [12], the Iterative Kth Ordered Scale Estimator (IKOSE) is proposed, which is very robust in estimating inlier scale. Based on IKOSE, we can distinguish inliers from outliers by using the estimated inlier scale. In [13], the Least Entropy-Like (LEL) estimator is proposed, which minimizes the entropy of the whole data but it does not differentiate inliers from outliers. Although it is computationally efficient in computing model parameters, the parameter estimate obtained by LEL is biased when outlier percentage is increased. Based on IKOSE and LEL, we propose an Adaptive Scale based Entropy-like Estimator (ASEE), which minimizes inliers' entropy. Compared with other robust estimators, ASEE is very simple and computationally efficient, and it can tolerate more than 90% outliers.

2. THE PROPOSED METHOD

The parametric model estimation problem can be described as follows: given a set of data points

$$X = [(x_1, y_1), \ldots, (x_N, y_N)]^t \in \mathrm{R}^{\mathrm{N}\times(\mathrm{d}+1)}$$

(with the explanatory variables $x_i = (x_{i1}, \ldots, x_{id}) \in \mathrm{R}^{\mathrm{d}}$, and the response variable $y_i \in \mathrm{R}^1$), estimate the regression coefficients of a parametric model $\tilde{\theta} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_d)^t \in \mathrm{R}^{\mathrm{d}}$ from X.

The classical linear regression model can be described as follows:

$$y_i = x_{i1}\theta_1 + \ldots + x_{id}\theta_d + e_i \quad (i = 1, \ldots, N). \tag{1}$$

The error term e_i is assumed to be normally distributed with $N(0, \sigma)$. Given an estimate $\tilde{\theta}$, the residual r_i corresponding to the i-th data point is estimated by using the following equation:

$$r_{i,\tilde{\theta}} = y_i - x_{i1}\tilde{\theta}_1 - \ldots - x_{id}\tilde{\theta}_d \tag{2}$$

* Corresponding author

The goal of a robust estimator is to robustly estimate the regression coefficients from input data which may involve a large number of outliers.

2.1. Scale Estimation

Given an estimated inlier scale $\tilde{s}$, inliers can be distinguished from outliers by satisfying the following equation:

$$|\frac{r_i}{\tilde{s}}| < \gamma \tag{3}$$

where γ is a constant factor and is often chosen as 2.5 to identify 98% of a normal distribution as inliers.

Many robust scale estimators have been proposed during the last decades. MEDian, MAD and KOSE [8] are the popular scale estimators. However, those scale estimators are often biased when outliers occupy the absolute majority of the whole data. Recently, an Iterative Kth Ordered Scale Estimator (IKOSE) [12] is proposed, which uses an iterative way to estimate inlier scale. It can be written as follows:

$$\tilde{s}_K \equiv |r_K| / \Phi^{-1}[0.5(1+\kappa')] \tag{4}$$

$$\kappa' \equiv K/n' \tag{5}$$

where $|r_K|$ is the Kth ordered absolute residual, $\Phi^{-1}[\cdot]$ is the augment of the normal cumulative density function, n' is the number of the inliers belonging to the model of interest and the K value is fixed to be 10% of the whole data points.

After the scale of inlier noise is obtained by IKOSE, we can distinguish inliers from outliers by using Eq. (3).

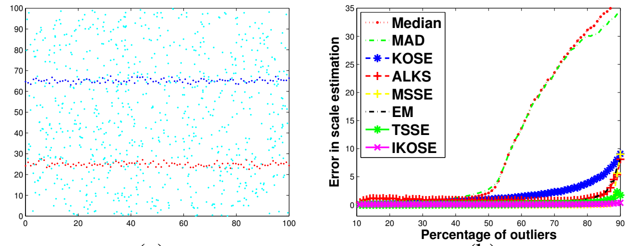
2.2. Performance of IKOSE in Scale Estimation

In this section, we compare the performance of eight scale estimators (MED, MAD, KOSE, ALKS, MSSE [14], EM [7], TSSE [10] and IKOSE [12]). The "parallel lines" data is generated with a total number of 1000 data points. Fig. 1(a) shows a snapshot of the data. Assuming that we know the true parameters of the model by which we can calculate the corresponding residuals and estimate the inlier scale by using the scale estimators. The true inlier scale of the data is set to 1.0. The data point number of the first line (in the red color) is gradually decreased from 900 to 100, at the mean time, the data point number of the second line (in the blue color) is fixed at 100, while the number of the gross noise increases. Thus, for the first line, the outlier percentage is increased from 10% to 90%.

We compute the error of the estimated inlier scale by using the following equation [12]:

$$\Upsilon(\tilde{s}, s_T) = max(\frac{\tilde{s}}{s_{\mathrm{T}}} - 1, \frac{s_{\mathrm{T}}}{\tilde{s}} - 1) \tag{6}$$

where s_{T} is the true inlier scale, and $\tilde{s}$ is the estimated inlier scale.



(a) (b)

Fig. 1: Comparison in scale estimation: (a) a snapshot of the "parallel lines" data with 90% outlier; (b) the error plots for the eight scale estimators.

	MED	MAD	KOSE	ALKS	MSSE	EM	TSSE	**IKOSE**
Mean	11.46	11.21	1.73	1.14	0.25	0.65	0.20	**0.11**
Std.Var.	13.07	12.44	1.92	0.98	1.07	1.31	0.33	**0.05**
Max.Err.	36.96	34.28	8.93	7.67	8.58	8.02	2.02	**0.35**

Table 1: Quantitative comparison of the competing methods to estimate inlier scale.

We repeat the experiments 60 times. Fig. 1(b) shows the averaged results. Table 1 shows the mean, the standard variance and the maximum errors of the estimated inlier scale. It can be seen from Fig. 1(b) and Table 1 that IKOSE achieves the best performance among the eight competing methods. Thus we choose IKOSE as the inlier scale estimator in our method.

2.3. ASEE Algorithm

Recently, the author of [13] developed a Least Entropy-Like (LEL) estimator. LEL is similar to RANSAC, M-estimators and ASSC. The aim of that method is to find a cost function to represent the scatter of the residuals (i.e., the entropy of the data). It is built on the concept of the (Gibbs) entropy [15]. LEL minimizes the entropy of the whole data and is very computationally efficient in some circumstances, but it does not distinguish the inliers from outliers. Moreover, the

Algorithm 1: The details of the ASEE algorithm

Input: A set of data points, the K value and the number of hypothese η
Output: The model parameters $\boldsymbol{\theta}_{\mathrm{ASEE}}$.
for $i = 1$ **to** η **do**
    Choose a minimal p-subset randomly.
    Estimate the model parameters $\boldsymbol{\theta}_i$ by using the p-subset.
    Calculate residuals $\Re'$.
    Estimate inlier scale by using **IKOSE**.
    Identify inliers by (3).
    Calculate $\mathbb{C}$, $\mathcal{C}_{i=1\ldots N}$, and $\boldsymbol{\Gamma}_i$ by (7) and (9).
    if $\boldsymbol{\Gamma}_i < \boldsymbol{\Gamma}_{min}$ **then**
        $\boldsymbol{\Gamma}_{min} = \boldsymbol{\Gamma}_i$, and $\boldsymbol{\theta}_{\mathrm{ASEE}} = \boldsymbol{\theta}_i$
    end
end

LEL estimator's penalty function is nonlinear and may have multiple local minima [13].

Based on IKOSE and LEL, we propose a novel robust estimator called as ASEE (Adaptive Scale based Entropy-like Estimator). It can identify inliers robustly by using the IKOSE algorithm as described in Sec. 2.1 and then minimizes the entropy of inliers while excluding outliers.

Having a model hypothesis, the residuals can be computed by (2). Let the sum of the squared residuals denote as: $\mathbb{C} = \sum_i^N r_i^2$. Then, the prior probability of r_i^2 can be written as:

$$\mathcal{C}_i = \frac{r_i^2}{\mathbb{C}}, \text{where } \mathcal{C}_i \in [0,1] \text{ and } \sum_{i=1}^{N} \mathcal{C}_i = 1. \quad (7)$$

The cost function of LEL is defined as follows [13]:

$$\mathbf{\Gamma}_{\text{LEL}} = \begin{cases} 0 & \text{if } \mathcal{C}_i = 0, \\ -\frac{1}{\log N} \sum_{i=1}^{N} \mathcal{C}_i \log \mathcal{C}_i & \text{otherwise.} \end{cases} \quad (8)$$

In LEL, it considers the whole data points in its objective function and finds the model parameters by minimizing $\mathbf{\Gamma}_{\text{LEL}}$. However, outliers may have some negative influence on the results of LEL. Intuitively, it is more robust to exclude outliers in the objective function which can improve the robustness of LEL in estimating model parameters. Thus, the objective function of ASEE is written as follows:

$$\mathbf{\Gamma}_{\text{ASEE}} = \begin{cases} 0 & \text{if } \mathcal{C}_i = 0, \\ -\frac{1}{N_k \log N_k} \sum_{i=1}^{N_k} \mathcal{C}_i \log \mathcal{C}_i & \text{otherwise.} \end{cases} \quad (9)$$

where N_k is the number of the inliers in data.

Compare (9) and (8), we can see that only inliers are considered in the proposed ASEE, by which the robustness of LEL is improved. Finally, ASEE can be written as follows:

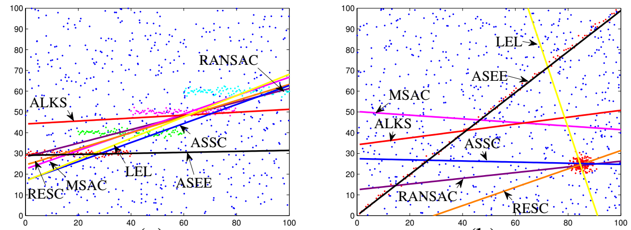
$$\boldsymbol{\theta}_{\text{ASEE}} = \arg\min_{\boldsymbol{\theta}} \mathbf{\Gamma}_{\text{ASEE}} \quad (10)$$

We use a "randomly sampling" scheme to choose the best hypothesis which yields the minimum score in (10), and sufficient hypotheses are required like many other robust estimators [6][10] for ASEE, so that at least one correct hypothesis is found.

The details of ASEE are described in Algorithm 1.

3. EXPERIMENTS

In this section, we evaluate the performance of ASEE on several synthetic and real image data. Firstly, we compare the performance of ASEE with those of several other robust estimators (RANSAC, MSAC, ALKS, RESC, ASSC, LEL) in line fitting by using synthetic data. Then we evaluate the performance of the competing methods by using three real image data.



(a) (b)

Fig. 2: Line fitting results: (a) the fitting results on the three-step signal with 85% outliers; (b) the fitting results on the signal involving clustered outliers and with 90% outliers.

	Error in A			Error in B		
	Mean	Std. Var.	Max. Err.	Mean	Std. Var.	Max. Err.
RANSAC	0.018	0.035	0.183	0.556	1.560	8.031
MSAC	0.253	0.189	0.427	3.814	2.798	6.645
ALKS	0.078	0.115	0.350	1.922	3.164	11.786
RESC	0.054	0.111	0.375	0.812	1.380	4.814
ASSC	0.044	0.112	0.463	0.811	1.886	7.181
LEL	0.407	0.093	0.512	7.726	2.258	12.064
ASEE	0.015	0.002	0.019	0.446	0.127	0.784

Table 2: Evaluation of the seven robust estimators for line fitting on the three-step signal.

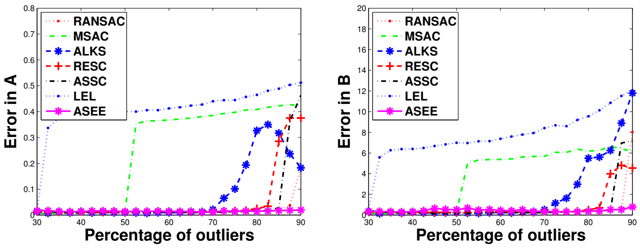
	Error in A			Error in B		
	Mean	Std. Var.	Max. Err.	Mean	Std. Var.	Max. Err.
RANSAC	0.078	0.243	0.986	2.074	6.023	23.915
MSAC	0.418	0.465	1.061	18.519	20.563	48.251
ALKS	0.099	0.250	0.882	4.064	9.800	34.759
RESC	0.234	0.711	2.908	17.892	54.557	228.064
ASSC	0.089	0.257	0.998	2.478	6.574	25.360
LEL	0.970	1.236	2.924	60.025	76.914	192.197
ASEE	0.006	0.001	0.009	0.326	0.062	0.531

Table 3: Evaluation of the seven robust estimators for line fitting on the signal with clustered outliers.

3.1. Synthetic Data

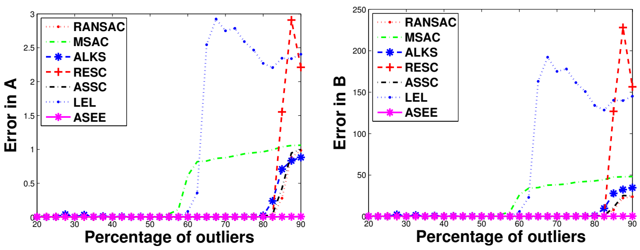
Firstly, we generate a three-step signal with a total of 1000 data points which are distributed in the range of [0 100]. The number of the data points belonging to the first line (in the red color in Fig. 2(a)) is gradually decreased from 700 to 100, the number of the data points belonging to the other three lines is fixed at 100, while the number of gross noise is increased from 0 to 600. Thus the outlier percentage for the first line is increased from 30% to 90%. The fitting results obtained by the competing methods are shown in Fig. 2(a). We repeat the experiments 60 times and show the mean errors in A and in B estimation (here, we use $y = Ax + B$ as the line model) at different outliers percentages in Fig. 3. We also show the mean, the standard variance and the maximum of the errors in A and B estimation in Table 2.

Secondly, we generate a signal with clustered outliers and gross outliers. The total data point number of the signal is 1000. The number of the data points belongs to the line in Fig. 2(b) is gradually decreased from 800 to 100. The number of the clustered outliers is fixed at 200 while the number of the gross outliers gradually increases. Thus, the outlier percentage to the line is increased from 20% to 90%. We repeat



(a) (b)

Fig. 3: The error plot for the three-step signal: (a) and (b) are the estimation errors in A and B vs. outlier percentage.



(a) (b)

Fig. 4: The error plot for the signal with clustered outliers: (a) and (b) are the estimation errors in A and B vs. outlier percentage.

the experiments 60 times and show the mean errors in A and B at different outlier percentages in Fig. 4. The mean, the standard variance and the maximum of the errors are shown in Table 3.

From the results, we can see that ASEE achieves the best performance among the seven robust estimators and it can tolerate up to 90% outliers. LEL begins to break down when the outlier percentage is up to 30% for the three-step signal and 60% for the signal with clustered outliers. The other estimators begin to break down as the outlier percentage increases.

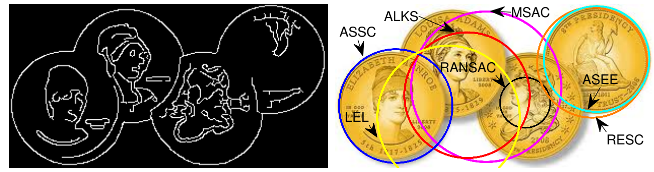
3.2. Real Image Data

In this section, we will give three examples to show the ability of ASEE in handling real image data.

The first example is to fit a coin[1] edge by the seven robust estimators. The edge image is obtained by using the Canny operator with a threshold 0.42. 3645 data points are obtained (as shown in Fig. 5(a)). The fitting results are shown in Fig. 5(b). ASEE fits the coin edge correctly, whereas RESC and ASSC fit the coin edge less accurately. The other estimators fail to correctly fit the edge of any coin.

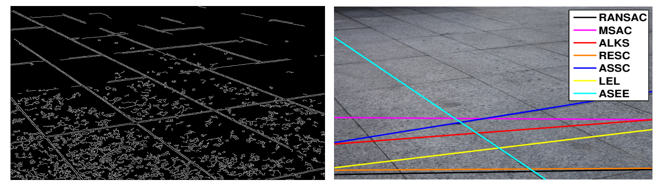
The second example is to fit a line in the pavement image, as shown in Fig. 6(a). The edge image is obtained by using the Canny operator with a threshold 0.29. 41783 data points are obtained. As Fig. 6(b) shows, ASEE gets the best results. In comparison, ASSC is a little biased while the other estimators completely fail.

[1]The image is taken from `http://www.coinlink.com`



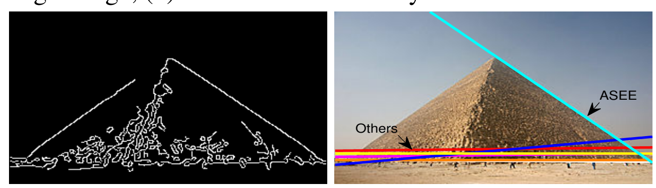
(a) (b)

Fig. 5: Fitting the edge of a coin: (a) the obtained edge image; (b) the results obtained by the seven estimators.



(a) (b)

Fig. 6: Fitting a line in the Pavement image: (a) the obtained edge image; (b) the results obtained by the seven estimators.



(a) (b)

Fig. 7: Fitting a line of the Khufu Pyramid: (a) the obtained edge image; (b) the results obtained by the seven estimators.

The third example is to fit the edge of the Pyramid in an image[2], as shown in Fig 7(a). We use the Canny operator with a threshold 0.25 to get the edge image. 2644 data points are obtained. The results are shown in Fig. 7(b). Only ASEE correctly fits the line model while the other estimators totally fail.

4. CONCLUSIONS

In this paper, we have proposed an efficient method, based on IKOSE and LEL, to estimate model parameters. Different to LEL, our method uses IKOSE to identify inliers and exclude outliers effectively, and then minimizes the entropy of the identified inliers. Compared with several other popular estimators, ASEE is very simple to implement, effective and efficient in practice, and more robust than other estimators, especially in dealing with extreme outliers.

5. ACKNOWLEDGEMENT

This work is supported by Natural Science Foundation of China (No. 61170179, No. 61202299 and No. 61201359), the Special Research Fund for the Doctoral Program of Higher Education of China (20110121110033), and the Xiamen Science & Technology Planning Project Fund (3502Z20116005) of China.

[2]The picture is taken from `http://en.wikipedia.org/wiki/Great_Pyramid_of_Giza`

6. REFERENCES

[1] S. Mittal, S. Anand, and P. Meer, "Generalized projection-based m-estimator," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 12, pp. 2351 –2364, dec. 2012.

[2] R. Hoseinnezhad and A. Bab-Hadiashar, "An m-estimator for high breakdown robust estimation in computer vision," *Computer Vision and Image Understanding*, vol. 115, no. 8, pp. 1145 – 1156, 2011.

[3] H. Liu and S. Yan, "Efficient structure detection via random consensus graph," in *IEEE Conference on Computer Vision and Pattern Recognition*, june 2012, pp. 574 –581.

[4] R. Raguram and J.-M. Frahm, "Recon: Scale-adaptive robust estimation via residual consensus," in *IEEE International Conference on Computer Vision*, nov. 2011, pp. 1299 –1306.

[5] P.J. Huber, *Robust Statistics*, New York, Wiley., 1981.

[6] M.A. Fischler and R.C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, June 1981.

[7] P. H. S. Torr and D. W. Murray, "The development and comparison of robust methods for estimating the fundamental matrix," *Int. J. Comput. Vision*, vol. 24, no. 3, pp. 271–300, Sept. 1997.

[8] K.-M. Lee, P. Meer, and R.-H. Park, "Robust adaptive segmentation of range images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 2, pp. 200 –205, feb 1998.

[9] X. Yu, T.D. Bui, and A. Krzyzak, "Robust estimation for range image segmentation and reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 530 –538, may 1994.

[10] H. Wang and D. Suter, "Robust adaptive-scale parametric model estimation for computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1459 – 1474, nov. 2004.

[11] H. Wang, D. Mirota, and G.D. Hager, "A generalized kernel consensus-based robust estimator," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 178 –184, jan. 2010.

[12] H. Wang, T.-J. Chin, and D. Suter, "Simultaneously fitting and segmenting multiple-structure data with outliers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1177–1192, June 2012.

[13] G. Indiveri, "An entropy-like estimator for robust parameter identification," *Entropy*, vol. 11, no. 4, pp. 560–585, 2009.

[14] A. Bab-Hadiashar and D. Suter, "Robust segmentation of visual data using ranked unbiased scale estimate," *Robotica*, vol. 17, no. 6, pp. 649–660, Nov. 1999.

[15] C.G. Chakrabarti and K. De, "Boltzmann-gibbs entropy: Axiomatic characterization and application.," *International Journal of Mathematics and Mathematical Sciences*, vol. 23, no. 4, pp. 243–251, 2000.