An Outlier Removal Method by Statistically Analyzing Hypotheses for Geometric Model Fitting

Yuewei Na, Guobao Xiao, and Hanzi Wang^(✉)

Fujian Key Laboratory of Sensing and Computing for Smart City,
School of Information Science and Engineering,
Xiamen University, Fujian 361005, China
na.yuewei@foxmail.com, x-gb@163.com,
wang.hanzi@gmail.com

Abstract. In this paper, we propose an outlier removal method which utilizes the information of hypotheses for model fitting. The proposed method statistically analyzes the properties of data points in two groups of hypotheses, i.e., “good hypotheses” and “bad hypotheses”. We show that the bad hypotheses, whose parameters are far from the parameters of model instances in data, also contain the correlation information between data points. The information can be used to effectively remove outliers from the data. Experimental results show the proposed method can effectively remove outliers on real datasets.

Keywords: Model fitting · Statistical analysis · Outlier removal

1 Introduction

Outliers exist in many datasets for geometric model fitting tasks. In order to alleviate the influence of outliers on model fitting, many robust model fitting methods like RANSAC [1], J-linkage [2], T-linkage [3], AKSWH [4], KF [5], PEARL [19] and EFM [20] have been developed. These methods sample minimal subsets (a minimal subset is the minimum number of data points which are required to compute the parameters of a geometric model) and generate a number of model hypotheses before they fit the models in data. Therefore, the quality of the generated hypotheses plays a critical role in model fitting methods, which boosts the development of guided sampling like [6, 7, 10, 12, 13, 16, 17].

The goal of guided sampling is to generate as large proportion of all-inlier hypotheses as possible. However, bad hypotheses, where the corresponding minimal subsets contain data points from different structures or the group of outliers, are often ignored by previous methods [2–4].

In this paper, we propose a novel outlier removal method, whose procedures are shown in Fig. 1. Based on the weighting method in [4], we propose a data point scoring strategy which utilizes the information of both bad hypotheses and good hypotheses. We regard hypotheses with high weights as “good hypotheses”. We note that the number of inliers belonging to good hypotheses are significantly larger than that of bad

hypotheses. Based on the observation, the scoring strategy essentially assigns a score for each data point by analyzing the frequency of two types of cases, i.e., a data point belongs to the minimal subset of a good hypothesis or a data point belongs to the minimal subset of a bad hypothesis. The former type of case decreases the score of the data point while the latter increases its score. To utilize the information of good hypotheses, the scoring strategy first assigns each data point a probability of being an outlier as shown in Fig. 1(b). For each data point, the probability of being an outlier is the exponential of the frequency of the former type of case with a base in interval $(0, 1)$. Then, in order to utilize the information of bad hypotheses, we introduce a two-dimensional “accumulation matrix”. If a pair of data points belong to the minimal subset of a bad hypothesis, the corresponding entries accumulate the minimum value of the probabilities of being outliers for this pair of data as shown in Fig. 1(e). The final score of each data point is given by the sum of the entries of the corresponding row in the accumulation matrix. Based on this scoring strategy, the score of an outlier will be larger than that of an inlier.

Fig. 1. An illustration of the proposed method for outlier removal on the fundamental matrix estimation problem. In (a) and (d), the “Cubebreadtoychips” image pair with 4 model instances of fundamental matrices is demonstrated. Correspondences are conducted by SIFT [8] matching. Gross outliers are labeled in red. The probability of being an outlier for each data point is given in (b). Outliers are on the left of the red line. In (e), we demonstrate the accumulation matrix, where the data points are grouped by real cluster membership. In (c) and (f), we show the remaining data points after removing outliers by the proposed method. (Color figure online)

The rest of the paper is organized as follows: We propose a novel outlier removal method in Sect. 2. We show the experimental results of the proposed method in Sect. 3 and draw the conclusions in Sect. 4.

2 The Proposed Outlier Removal Method

Given an input dataset $\mathcal{X} = \{x_i\}_{i=1}^N$ which contains N data points, the proposed method first samples M hypotheses by Multi-GS [7]. Denote the whole set of hypotheses by $\theta = \{\theta_j\}_{j=1}^M$. We assign each hypothesis θ_j a weight w_j and separate all hypotheses into two groups, i.e., good hypotheses and bad hypotheses, based on their weights (described in Sect. 2.1). The good hypotheses are used for computing the probabilities of being outliers (described in Sect. 2.2). The bad hypotheses are used for accumulate the probabilities of being outliers (described in Sect. 2.3). Finally, we compute the score of each data point and dichotomize inliers and outliers (described in Sect. 2.4).

2.1 Dichotomizing Good and Bad Hypotheses

Effectively weighting hypotheses plays an important role in many robust estimators. One simple and natural way was proposed in RCD [9] which was used to deal with the robust circle detection problem. RCD weights each hypothesis by counting the number of data points whose residuals are smaller than a manually set inlier scale. However, the method is sensitive to the inlier scale. In addition, the difference of weights between a good hypothesis and a bad one is not significant.

In [4], the authors propose a novel method to weight hypotheses. That is, given a set of data points $\mathcal{X}$ and a set of hypotheses θ , the residual of data point x_i to the parameter of hypothesis θ_j is given by

$$r_i^j = F(x_i, \theta_j) \quad (1)$$

where $F(\cdot)$ is a function that computes the residual based on the arguments x_i and θ_j as in [4]. Denote the residual set of the hypothesis θ_j by $R_j = \{r_i^j\}_{i=1}^N$ and the sorted residuals in an ascending order by $\hat{R}_j = |\hat{r}_i^j|$, then the inlier scale estimated by IKOSE [4] can be written as follows:

$$S_K^j = \hat{R}_j^K / \psi^{-1}\left(\frac{1}{2}(1 + \kappa)\right) \quad (2)$$

$$\kappa = K / \tilde{n} \quad (3)$$

where $\hat{R}_j^K$ is the K th ordered absolute residual; $\psi^{-1}(\cdot)$ is the argument of the normal cumulative density function; $\tilde{n}$ is the number of inliers of the model θ_j , and K is fixed to 10 % of the whole data points in our experiments. The weight of hypothesis θ_j can be calculated by [4]:

$$w_j = \frac{1}{N} \sum_{i=1}^N \frac{KN(r_i^j/h(\theta_j))}{S_K^j h(\theta_j)} \quad (4)$$

where $h(\cdot)$ and $KN(\cdot)$ are the bandwidth and the kernel, respectively.

In order to obtain good hypotheses from the whole hypotheses set, we apply the EM algorithm [14] to fit a Gaussian mixture model. The model preserves more number of good hypotheses comparing to the entropy thresholding method used in [4] and other thresholding methods such as histogram shape-based methods in [15].

2.2 Calculating Probabilities of Being Outliers

In this section, we calculate the probability of being an outlier for each data point based on the information of good hypotheses. Denote that the event “Data point x_i belongs to the minimal subset of a good hypothesis” as E_g . The proposed method is based on the observation that the frequency of the event E_g based on inliers is much higher than the frequency based on outliers. As hypotheses are dichotomized in Sect. 2.1, we denote the minimal subsets of good hypotheses and the minimal subsets of bad hypotheses as $\mathcal{G} = \{g_l\}_{l=1}^{M_g}$ and $\mathcal{B} = \{b_l\}_{l=1}^{M_b}$, respectively, where $M_g + M_b = M$, based on the dichotomy. The frequency f_i of E_g for data point x_i is given by the number of l s where $x_i \in g_l$ (l is the index of the minimal subsets). After calculating the frequencies, the probability of being an outlier for data point x_i is given by:

$$\text{Prob}(x_i) = C^{f_i} \quad (5)$$

where C is a constant in the open interval $(0, 1)$ in order to assure that the probability $\text{Prob}(x_i)$ takes values in the close interval $[0, 1]$. As shown in Fig. 2, the difference between the probabilities of being outliers for inlier data and the probabilities of being outliers for outlier data becomes significant, as the number of good hypotheses grows.

Fig. 2. Changes of the probabilities of being outliers as the number of good hypotheses grows for the “Cubebreadtoychips” dataset in Fig. 1. The good hypotheses number used to calculate the probabilities are respectively 1000, 2000 and 3000 in (a), (b) and (c). The constant C is set to 0.99. Data points on the left of the red line correspond to the outliers. (Color figure online)

2.3 Accumulating Probabilities of Being Outliers

To utilize the information of bad hypotheses, we denote that the event “Data point x_i belongs to the minimal subset of a bad hypothesis” as E_b . Our method for accumulating probabilities of being outliers is based on the observation on the frequency of the event E_b . We observe that the frequency of event E_b based on outliers is much higher than that based on inliers. In our method, we introduce an N -by- N accumulation matrix S , each of whose entry is initialized with zero. For each pair of data points $\{x_i, x_j\} \subseteq b_t (i, j = 1..N, t = 1..M_b)$, we accumulate the probabilities of being outliers by the following equation:

$$S(i, j) = S(i, j) + \min(\text{Prob}(x_i), \text{Prob}(x_j)) \quad (6)$$

We can see that the entry $S(i, j)$ will be significantly large if data points x_i and x_j are outliers, since the probabilities of being outliers for outlier data are larger than that of inliers as mentioned in Sect. 2.2. The accumulation process is shown in Fig. 3. We can see that the differences between entries become larger as the number of bad hypotheses, which are used to accumulate probabilities of being outliers, increases. Note that the entries in the upper-left submatrix are significantly larger than the other entries. This phenomenon coincides with our analysis on entry $S(i, j)$ when both data points x_i and x_j are outliers.

Fig. 3. Evolution of the accumulation matrix as the number of analyzed bad hypotheses grows for the “Cubebreadtoychips” dataset in Fig. 1. The number of bad hypotheses used to calculate the accumulation matrix are respectively 5000, 10000 and 30000 in (a), (b) and (c). The ranges of the grayscale images are based on the minimum and maximum values in the corresponding accumulation matrix. The data points are grouped based on the true cluster membership.

2.4 The Complete Outlier Removal Method

After analyzing all the generated hypotheses, the score of a data point x_i is given by:

$$\text{Score}(i) = \sum_{j=1}^N S(i, j) \quad (7)$$

By Eq. (6), the score of an outlier will be significantly larger than that of an inlier. As shown in Fig. 4(a), the histogram of the scores has two significantly separated groups. Note that the distribution of scores follows 1D two-component Gaussian mixture model:

Fig. 4. Illustration of the clustering procedure for the “Cubebreadtoychips” dataset in Fig. 1. The histogram of data point scores is shown in (a). The red line corresponds to the inlier/outlier dichotomization threshold computed by the EM algorithm. The corresponding probability density function $p(x)$ is presented in (b). The scores are computed based on the accumulation matrix in Fig. 1(e). (Color figure online)

$$p(x) = \sum_{i=1}^2 \pi_i \mathcal{N}(x | \mu_i, \sigma_i) \quad (8)$$

where π_i is the component weight and $\mathcal{N}$ is a Gaussian with mean μ_i and standard deviation σ_i . Then the clustering membership of each data point can be explored by using the EM algorithm. The cluster of data points with a higher mean score will be classified as outliers and removed. Figure 4(b) presents the probability density function of the Gaussian mixture model, which almost perfectly fits the distribution of the histogram. Algorithm 1 details the complete process of the proposed outlier removal method.

Algorithm 1. The proposed outlier removal method

Input: data points set $\mathcal{X}$ containing N points, constant C for computing probabilities of being outliers and K value for IKOSE

Output: inlier/outlier dichotomization result

- 1: Generate an initial hypothesis set θ by Multi-GS.
 - 2: Estimate the inlier scales of the hypotheses by IKOSE.
 - 3: Calculate the weights of each hypothesis by Eq. (4).
 - 4: Dichotomize good and bad hypotheses by the EM algorithm.
 - 5: Calculate the probability of being an outlier $\text{Prob}(x_i)$ for each point x_i .
 - 6: Accumulate probabilities of being outliers in S .
 - 7: Calculate the scores of data points by Eq. (8).
-

3 Experimental Results

In this section, we compare the proposed method with the outlier removal method introduced in KF [5] on AdelaideRMF [11].¹ For fairness, the input arguments of both methods, such as C for computing probabilities of being outliers in our method and the step size in KF, are tuned separately on each dataset for the best performance. In order to measure the outlier removal performance on model fitting problems, we introduce the definition of *misclassification error* (ME) similar to [3]:

$$\text{ME} := \frac{\#\text{Misclassified Points}}{\#\text{Total Points}} \times 100\% \quad (9)$$

where a point is “misclassified” if an inlier is classified as an outlier according to the ground truth and vice versa.

3.1 Outlier Removal for Homography Estimation

In this section, we compare the proposed method with KF for the homography estimation problem. Datasets in [11] contain sets of correspondences described by several homographies. We use Direct Linear Transformation (4 points required for each minimal subset) as the hypothesis generation method and Sampson distance [18] as the method of computing residuals. Table 1 shows the mean and median ME on various datasets over 50 repetitions. As shown in Table 1, the experimental results show that the proposed method can effectively identify outliers in the data. Figures 5(a) and (b) shows the comparison on the data “Johnsona” and “Napierb”. These two figures clearly shows that the proposed method not only effectively removes more outliers, but also preserves more inliers compared with KF. Note that KF can correctly identify all the outliers in the “Oldclassicswing” data. This is because the inlier scale and outlier percentage of the data is relatively small, compared with the other datasets.

3.2 Outlier Removal for Fundamental Matrix Estimation

We also compare the performances of the two competing methods for fundamental matrix estimation task. Table 2 shows the mean and median ME obtained by the two competing methods on different datasets over 50 repetitions. It can be seen from Table 2 that the proposed method outperforms KF on average on the fundamental matrix estimation task. Figures 5(c) and (d) demonstrate the performance of the proposed method and KF on the data “Breadtoycar” and “Breadcubechips”. These two subfigures show that the proposed method can more effectively dichotomize inliers and outliers for the fundamental matrix estimation task.

¹ <http://cs.adelaide.edu.au/~hwong/doku.php?id=data>

Table 1. The mean and median results over 50 repetitions for homography etimation (the best results are marked in bold).

Dataset		KF	Our method
Johnsona	mean	1.53 %	1.38 %
	median	1.42 %	1.34 %
Nese	mean	1.18 %	0.81 %
	median	1.18 %	0.39 %
Oldclassicswing	mean	0 %	1.56 %
	median	0 %	1.29 %
Physics	mean	11.21 %	3.92 %
	median	11.32 %	2.83 %
Sene	mean	6 %	4.89 %
	median	6 %	3.60 %
Napiera	mean	21.85 %	11.26 %
	median	21.85 %	9.6 %
Napierb	mean	6.74 %	2.92 %
	median	6.56 %	2.70 %

Fig. 5. Experimental results obtained by the proposed method and KF. Top row: The ground truth. Middle Row: The outlier removal performance obtained by KF. Bottom Row: The outlier removal performance obtained by the proposed method. Outliers are labeled in red in all images. We show two results selected from Tables 1 and 2 for each kind of model fitting problem. (Color figure online)

Table 2. The mean and median results over 50 repetitions for fundamental matrix estimation (the best results are marked in bold).

Dataset		KF	Our method
Breadcartoychips	mean	3.49 %	3.23 %
	median	3.75 %	2.95 %
Breadcube chips	mean	7.39 %	4.03 %
	median	7.39 %	2.17 %
Breadtoycar	mean	9.64 %	5.92 %
	median	9.64 %	3.01 %
Carchipscube	mean	3.64 %	3.85 %
	median	3.64 %	4.85 %
Cubebreadtoychips	mean	8.92 %	1.30 %
	median	8.87 %	1.22 %

4 Conclusions

In this paper, we propose a novel outlier removal which utilizes the useful information of both good and bad hypotheses. We argue that bad hypotheses also contain information about the correlations between data points as good hypotheses do. The experimental results show that the proposed method is overall a better outlier removal comparing to KF.

Acknowledgment. This work was supported by the National Natural Science Foundation of China under Grants 61472334 and 61170179, and supported by the Fundamental Research Funds for the Central Universities under Grant 20720130720.

References

1. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**(6), 381–395 (1981)
2. Toldo, R., Fusiello, A.: Robust Multiple Structures Estimation with J-Linkage. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part I*. LNCS, vol. 5302, pp. 537–547. Springer, Heidelberg (2008)
3. Magri, L., Fusiello, A.: T-Linkage: a continuous relaxation of j-linkage for multi-model fitting. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3954–3961 (2014)
4. Wang, H., Chin, T.J., Suter, D.: Simultaneously fitting and segmenting multiple-structure data with outliers. *pattern analysis and machine intelligence. IEEE Trans.* **34**(6), 1177–1192 (2012)
5. Chin, T.J., Wang, H., Suter, D.: Robust fitting of multiple structures: the statistical learning approach. In: *Proceedings of IEEE International Conference on Computer Vision*, pp. 413–420. IEEE (2009)

6. Chum, O., Matas, J.: Matching with PROSAC-progressive sample consensus. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 220–226. IEEE (2005)
7. Chin, T.J., Yu, J., Suter, D.: Accelerated hypothesis generation for multistructure data via preference analysis. *pattern analysis and machine intelligence. IEEE Trans.* **34**(4), 625–638 (2012)
8. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
9. Chen, T.C., Chung, K.L.: An efficient randomized algorithm for detecting circles. *Comput. Vis. Image Underst.* **83**(2), 172–191 (2001)
10. Pham, T.T., Chin, T.J., Yu, J., Suter, D.: The random cluster model for robust geometric fitting. *Pattern analysis and machine intelligence. IEEE Trans.* **36**(8), 1658–1671 (2014)
11. Wong, H.S., Chin, T.J., Yu, J., Suter, D.: Dynamic and hierarchical multi-structure geometric model fitting. In: IEEE International Conference on Computer Vision (ICCV), 2011, pp. 1044–1051. IEEE (2011)
12. Kanazawa, Y., Kawakami, H.: Detection of planar regions with uncalibrated stereo using distributions of feature points. In: BMVC, pp. 1–10 (2004)
13. Tordoff, B.J., Murray, D.W.: Guided-MLESAC: faster image transform estimation by using matching priors. *pattern analysis and machine intelligence. IEEE Trans.* **27**(10), 1523–1535 (2005)
14. Bishop, C.M.: *Pattern Recognition And Machine Learning*. Springer, New York (2006)
15. Sezgin, M.: Survey over image thresholding techniques and quantitative performance evaluation. *J. Electron. Imaging* **13**(1), 146–168 (2004)
16. Chum, O., Matas, J., Kittler, J.: Locally optimized RANSAC. In: Michaelis, B., Krell, G. (eds.) DAGM 2003. LNCS, vol. 2781, pp. 236–243. Springer, Heidelberg (2003)
17. Wong, H.S., Chin, T.J., Yu, J., Suter, D.: Efficient multi-structure robust fitting with incremental top-k lists comparison. In: Asian Conference on Computer Vision–ACCV 2010, pp. 553–564. Springer, Berlin Heidelberg (2011)
18. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York (2003)
19. Isack, H., Boykov, Y.: Energy-based geometric multi-model fitting. *Int. J. Comput. Vis.* **97**(2), 123–147 (2012)
20. Hossam, N.I., Yuri, B.: Energy based multi-model fitting & matching for 3D reconstruction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1146–1153 (2014)