International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2017; 2(2): 31-34

www.mathsjournal.com
Received: 15-01-2017
Accepted: 16-02-2017

Ravi. J
Associate Professor, Department of Statistics, Vivekanandha College for Women, Unjanai, Tiruchengode, Namakkal, Tamil Nadu, India

Correspondence
Ravi. J
Associate Professor, Department of Statistics, Vivekanandha College for Women, Unjanai, Tiruchengode, Namakkal, Tamil Nadu, India

A robust measure of pairwise distance estimation approach: RD-RANSAC

Ravi J

Abstract
A method for detection of outliers is proposed which a Robust Distance to be used RANSAC (RD-RANSAC). A novel idea on hot to make RANSAC repeatable is presented, which will find the optimum set in nearly run for multi-model. Robust methods are capable of discriminating correspondence outliers, thus, obtaining better results. Our proposed method is an improvement of RANSAC which takes into account additional information of the quality of the matches to largely reduce the computational cost of the pair wise distance estimation by Rousseeuw's Minimum Covariance Determinant (MCD). However, even in quite large samples, the chi-square approximation to the distance of the sample data from the MCD center with respect to the MCD shape is poor [1]. RANSAC can only estimate one model for a particular data set. The two or more model exists; RANSAC may fail to find either one. The problem is hard as the number of outlier is usually large, possibly larger than 50%, thus powerful estimation techniques are need. Experiments with up to 80% outlier prove the efficiency of RANSAC [2]. RANSAC is not always able to find the optimum set even for moderately contaminated sets and it usually performs badly when the number of inliers is less. However this work proposes a new robust method for pairwise distance estimation to combine the benefits of RANSAC algorithm, namely improved quality, reduced computational time and less parameter to adjust and powerful estimation techniques up to more than 80% outlier prove the efficiency.

Keywords: RANSAC, robust statistics, MCD, RD-RANSAC.

1. Introduction
Many computer vision approaches for extracting features and inferring image content are based on detecting some point of interest. It is difficult using RANSAC while trying to run tests of other parameters involved in the application, as the set of inliers for the same pair of images may vary in each run. For medical applications such as the one described later it is important that the result does not differ if the application is run more than once. Furthermore, standard RANSAC does not try to find the optimal set of inliers, instead it stops when the probability of findingmore inliers reaches its predefined threshold. The consequence is that it does not perform well for highly contaminated sets. RANSAC is aimed to determine function parameters when there exist gross -erroneous samplesthat can mislead the parameter estimation. Plane extraction methods in the literature may be broadly classified into two categories: Region-growing and clustering algorithms and model fitting methods. A region going algorithm [3] select some seed points in range data and grown them into regions based on the homogeneity of local features.

Robust statistical methods were first adopted in computer vision to improve the performance of feature extraction algorithms at the bottom level of the vision hierarchy. This method tolerates the presence of data points that do not obey the assumed model such points are typically called "outlier" [4]. Recently, various robust statistical methods have been developed and applied to computer vision tasks such as the standard RANSAC is not always able to find the optimal set even for moderately contaminated sets and it usually performs badly when the number of inliers is less than 50% [5] in this method is based on several known method, which we modify in a unique way and together they produce a result. RANSAC and related hypothesis size-and-verify methods [6-10] have been applied to many vision problems. The RANSAC plane fitting produces inlier patches (in the tread, riser and stair wall planes) that connect together to form a plane [11]. The method uses RANSAC to generate the so-called preferences sets for all data points, form which multiple models are extracted by an agglomerative clustering algorithm, called J-linkage. In this paper, we propose an improved method, called a robust measure of pairwise distance estimation approach using Mahalanobis Squared Distance (or MSD) measurement method, for extraction this plane in RANSAC. The remainder of the paper is organized as follows. Section 2 details the introduction and disadvantages of RANSAC. Section 3 presents the RD-RANSAC. Section 4 presents experimental results with real sensor data and the paper is concluded in section 5.

2. Random Sample Consensus (RANSAC) techniques

Martin A. Fischler and Robert C. Bolles (1981) was proposed Random Sample Consensus [12] algorithm for model fitting with applications of image analysis and other computer vision tasks. The RANSAC proceeds as follows: Repeatedly, subsets of the input data are randomly selected with replacement, and model parameters fitting these subsets are computed. Then the quality of parameters is evaluated on the input data. Different cost functions have been proposed, the standard being the number of data points consistent with model. The process is terminated when the probability of finding a better model become lower than a user specified probabilityη_0. The 1-η_0 confidence in the solution holds for all levels of contamination of the input data, that is, for any number of outlier within the input data. The speed of RANSAC algorithm depends on two factors: The number of random samples and the number of the input data points. In all common settings where RANSAC is applied, almost all models whose quality is verified are incorrect with arbitrary parameters originating from contaminated samples. Such models are consistent with only a small number of the data points.

RANSAC Motivated for computer vision. The gross errors (outlier) spoil estimation and localization algorithms in computer vision do have gross error. In difficult problems the portion of good data may be even less than ½. But standard robust estimation techniques hardly applicable to data with less than ½ good. But RANSAC can only estimate one model for a particular data set. As for any one-model approach when two (or more) model instances exist, RANSAC may fail to find either one. RANSAC is the measured data has total of N samples with unknown fraction of inliers γ. To estimate true model parameters we would like to label data as outliers and inliers and estimate the model parameters from inliers only. S this labeling is initially unknown RANSAC tries to find outlier-free data subset randomly, in several attempts. To maximize the probability of selecting sample without outliers RANSAC tests only samples of minimal size.

RANSAC distance (d*) in image Metrics (Normalized) Sum of Squared Differences as follows,

$d^* = \arg\min_d \|w_L - w_R(d)\|^2 = \mathrm{Argmax}_d w_L w_R(d)$

where$w_L\, w_R(d) = \cos\theta$

$d^2 = d(X, H^{-1}X')^2 + d(X', HX)^2$

The RANSAC algorithm consists of M iteration of the following steps:

K:=0, Repeat until P(Better solution exists)<η (a function of C* and number of steps k)

K;=k+1

Steps:

(1) Select randomly set $S_k \subset U$, $|S_k|$=s
(2) Compute parameter θ_k=f(S_k)
(3) Compute cost $C_k = \sum_{x\in U} p(\theta_k, x)$
(4) if $C^* < C_k$ then $C^* := C_k$, $\theta^* = \theta_k$

end.

3. Robust Distance Distribution

Various methods for detecting outliers have been studied (Atkinson 1994; Barnett and Lewis 1994, Becker and Gather 1999, 2001; Davies and Gather 1993; Gatherand Becker 1997; Gnanadesikan and Kettenring 1972; Hadi 1992, 1994; Hawkins 1980; Maronna and Yohai 1995; Penny 1995; Rocke and Woodruff 1996; Rousseeuw and van Zomeren 1990). One way to identify possible Outliers is to calculate a distancefrom each point to a "center" of the data. An outlier would then be a point with a distance larger than some predetermined cutoff. A Conventional measurement of quadratic distancefrom a point *X* to a location *Y* given a shape *S*, in the setting is

$$d^2_{MSD}(x,y) = \left[\begin{pmatrix} x_i \\ y_i \end{pmatrix} - \begin{pmatrix} \bar{X}_{MSD} \\ \bar{Y}_{MSD} \end{pmatrix}\right]^T - S^{-1}_{MCD}\left[\begin{pmatrix} x_i \\ y_i \end{pmatrix} - \begin{pmatrix} \bar{X}_{MSD} \\ \bar{Y}_{MSD} \end{pmatrix}\right]^1 \quad (1.1)$$

This quadratic form is often called the Mahalanobis squared distance (MSD). If there are only a few outliers, large values of $d^2_{MSD}$and covariance matrix, indicate that the point xi is an outlier (Barnett and Lewis 1994). The distribution of the MSD with both the true location and shape parameters and the standard sample location and shape parameters is well known (Gnanadesikan and Kettenring 1972). However, the standard sample location and shape parameters are not robust to outliers, and the distributional fit to the distance breaks down when robust measures of location and shape are used in the MSD (Rousseeuw and van Zomeren 1991). Determining exact cutoff Values for outlying distances continue to be a difficult problem.

In trying to detect a single outlier in a multivariate normal sample, d^2_{MSD}willidentify a sufficiently outlying point. In data with clusters of outliers, however, the distance measure d^2_{MSD} breaks down (Rocke and Woodruff 1996). Datasets with multiple outliers or clusters of outliers are subject to problems of masking and swamping (Pearson and Chandra Sekar 1936). Problems of masking and swamping can be resolved by using robust estimates of shape and location, which by definition are less affected by outliers. Outlying points are less likely to enter into the calculation of the robust statistics, so they will be less likely to influence the parameters used in the MSD. The inlying points, which all come from the underlying distribution, will completely determine the estimate of the location and shape of the data. We use Rousseeuw's minimum covariance determinant (MCD) (Rousseeuw 1985) to estimate the location and shape of the data. When using the MCD in the distance function, however, we no longer have well-known distributional information for the distances. Using the motivation that independent data distances have an *F* distribution, we apply an adjusted *F* distribution to the extreme sample points. The *F* distribution is more representative of the extreme points than the more commonly used $\chi 2$ distribution [1].

Our pair wise distances are calculated using the Mahalanobis Squared Distance measurement. Because RD-RANSAC is able to automatically set the threshold to determine inlier while at the same time producing slightly better accuracy than

RANSAC. We propose in this document the combination of the RD-RANSAC and RANSAC algorithms to get the bestof both algorithm in terms of speed and

Quality of the estimation. We call this algorithm RD-RANSAC behaves exactly as RANSAC except that it picks distance.

Where MCD=$(\overline{X^*}_J S^*_j)$; J=(set of h points:$|S^*_J|$= $|S^*_K|$ for all sets $|K|$=h)

$$\bar{X}^*_{MCD} = \frac{1}{h}\sum_{i\epsilon j} x_i \quad (1.2)$$

$$S^*_{MCD} = \frac{1}{h}\sum_{i\epsilon j}\left(x_i - \overline{X^*}_J\right)\left(x_i - \overline{X^*}_J\right)^T \quad (1.3)$$

h= $\left[\frac{n+p+1}{2}\right]$, n=17 ie., The number of arrays in our data, and p=2. The value h is the highest possible breakdown point for the MCD, ie., it is minimum number of points which must not be outlying.

3.1 Cutoffs to be finding

After find d^2_{MSD}, for each pair wise distance, we need a mechanism for evaluating whether the distance is sufficient outlying with respect to the relationship between the two genes.

MSD distance with MCD shape and location parameters are known to be robust with an F-distribution when the data are normally distributed.

We find the cutoff values using an F-distribution to choose which points are outliers. We use the similarity equation.

$\frac{c(m-p+1)}{pm} d^2{}_s(X_I - \bar{X})(X_I - \bar{X})$ from equation (4), we need estimates the c and m. to find an estimate c. $\hat{c}$, and an estimate m, $\hat{m}$, we use the set of equations below. $\hat{c}$,is solved for in equation (1.4), and we use equation (6) to solve for$\hat{m}$.

$$\hat{c} = \frac{P(\chi_{p+2} < \chi_{p,h/n})}{h/n} \quad (1.4)$$

$$\alpha = \frac{n-h}{n}, \qquad \alpha = P\left(\chi^2{}_\alpha < q_\alpha\right), \qquad C_\alpha = \frac{1-\alpha}{P(\chi^2{}_\alpha < q_\alpha)}$$

$$C_2 = \frac{-P\left(\chi^2{}_{p+2} < q_\alpha\right)}{2}, \qquad C_3 = \frac{-P\left(\chi^2{}_{p+4} < q_\alpha\right)}{2}, \qquad C_4 = 3*C_3$$

$$b_1 = \frac{C_{\alpha(C_3 - C_4)}}{1-\alpha}, \qquad b_2 = 0.5 + \frac{C_\alpha}{1-\alpha}\left(C_3 - \frac{q_\alpha}{p}\left(C_2 - \frac{1-\alpha}{2}\right)\right) \quad (1.5)$$

$$v_1 = (1-\alpha)b_1{}^2\left\{\left(\frac{C_\alpha q_\alpha}{p} - 1\right)^2 - 1\right\} - 2C_3 C_\alpha{}^2\left\{\begin{matrix} 3(b_1 - pb_2)^2 \\ +(p+2)b_2(2b_1 - pb_2) \end{matrix}\right\}$$

$$v_2 = n\{b_1(b_1 - pb_2)(1-\alpha)\}^2 C^2{}_\alpha$$

$$v = \frac{v_1}{v_2}, \quad \hat{m} = C^2{}_\alpha$$

3.2 Robust Distance-RANSAC (RD-RANSAC)

The robust distance is to be applied in RANSAC it will be get accuracy of outlier to be identified because of F estimate. The RANSAC algorithm is robust but it is not completely free from the effect of outliers and is slow for large datasets. Recently introduced high breakdown and fast Minimum Covariance Determinants (MCD) based robust distance to identify outliers and a MCD based robust approach is used as outlier resistant techniques for plane fitting [13]. The idea of RD-RANSAC algorithm is to repeatedly select a random subset S of the data, to determine a solution P=F(S) and evaluate it with other data.

Procedure

- Determine, n,k,t and d
- To assert a model fits well until k iterations have occurred and Draw a sample of n points from the data uniformly and at random
- Fit to that set of n points For each data point outside the sample
- Test the Mahalanobis distance from the point to the line against t if the Mehalanobis distance from point to the line is less than t, the point is close
- end
- If there are d^2_{MSD} (x,y) or more points close to the line then there is a good fit.
- Refit the line using all these points
- end
- Use the best fit (F- distribution) from this collection using the fitting error as criterion.

4. Experimental Results

We developed simulation software to investigate the RD-RANSAC. The user specifies the number of points, the percentage of outliers, the straight line and the measuring precision used for generating the inlier data using TLS model. In addition, the user specifies the required minimum probability for success and the expected error rate, independently of the generation in order to investigate the effect of erroneous assumptions

In this section presents the simulation study with result to compare the performance of RD-RANSAC method with RANSAC, NAPSAC and INAPSAC methods. This simulation study is carried out for the different number of threshold such as 2, 4 and 6 and for various samples sizes, n=100, n=500 and n=1000. The data is generated using TLS model. The number of inliers and the corresponding mean value for those inliers are estimated using various RANSAC techniques, the results are summarized in the below table.

Table 1: The Estimated Result of RD-RANSAC with other RANSAC techniques

T	Methods	n=100		N=500		N=1000	
		Mean	Inliers	Mean	Inliers	Mean	Inliers
2	TLS	0.074522	95	0.022838	475	0.027711	950
	RANSAC	0.085800	93	0.031563	457	0.021601	916
	NAPSAC	-0.490461	85	0.504939	441	0.350763	922
	INAPSAC	0.500666	95	-0.263359	460	0.017260	936
	RD-RANSAC	0.522222	95	0.452365	473	0.325487	947
4	TLS	0.175107	95	-0.049626	475	-0.033793	950
	RANSAC	-0.603010	91	0.269810	47	0.189419	939
	NAPSAC	-0.442057	84	0.112394	471	1.063397	931
	INAPSAC	0.314658	92	-0.238840	470	0.819703	937
	RD-RANSAC	0.658954	93	-0.658945	472	0.547896	945
6	TLS	-0.024324	95	-0.017501	475	0.051102	950
	RANSAC	-0.029559	90	1.252636	483	0.953964	962
	NAPSAC	0.121398	90	0.200737	483	-0.381640	963
	INAPSAC	-0.075238	89	-0.038167	485	-1.446939	966
	RD-RANSAC	0.125489	93	0.356898	472	-0.121245	944

It is observed form the above table, in all the situations the RD-RANSAC techniques gives the better result than the other SAC procedures, the number of inlier points are more than the other method.

5. Conclusions

This paper demonstrates the ability of new Robust Distance estimator and the data generated by used TLS model. Robust Distance -Random sample consensus (RD-RANSAC) is an effective tool to detect the true parameters also in presence of highly corrupted data. According [1] the minimum covariance determinant (MCD) is a robust estimator with a high breakdown. However, even in quite large samples, the chi-squared approximation to the distances of the sample data from the MCD center with respect to the MCD shape ispoor. So we provide an approximation that gives accurate outlier rejection points for various sample sizes.

Altogether, RD-RANSAC is a powerful tool to cope with large percentages of blunders. It can be successfully used in digital image analysis, feature matching procedures as well as for automatic relative orientation of images to detect corners perfectly. RD-RANSAC has been used extensively and has proven to give better performance than other SAC procedures.

References

1. Hardin J, Rocke DM. The Distribution of Robust Distance. Journal of Computational and Graphical Statistics, 14(4):1-19. DOI: 10.1198/106186005X77685.
2. Ruzgiene B, Forstner W. RANSAC For Outliers Detection”, Geodesy and Cartography, 2005; 32(3):83-87.
3. Kaushik R, Xiao J, Joseph S, Morri W. Fast Planner Clustering and polygon Extraction from Noisy Range Image Acquired in Indoor Environments, in proc. IEEE International conference on Mechatronics and Automation, 2010, 483-488.
4. Radha M, Muthukrishnan R. INAPSAC: A New Robust Inlier Identification Technique, Journal of Advanced Computer Science and Technology, 2012; 1(4):284-290.
5. Hast A, Nysjo J, Marchetti A. Optimal RANSAC-Towards a Repeatable Algorithm for Finding the Optimal Set, tenure of an ERCIM Alain Bensuosan Fellowship Programme.
6. Chum O, Matas J, Kittler J. Locally Optimized RANSAC, Proc. Ann. Symp. German Assoc. for Pattern Recognition (DAGM ’03), 2003.
7. Chum O, Werner T, Matas J. Epipolar Geometry Estimationvia RANSAC Benefits from the Oriented Epipolar Constraint,”Proc. Int’l Conf. Pattern Recognition (ICPR ’04), 2004; 1:112-115.
8. Fischler M, Bolles R. Random Sample Consensus: AParadigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Comm. ACM, 1981; 24(6):381-395.
9. Nister D. Preemptive RANSAC for Live Structure and Motion Estimation,” Proc. Int’l Conf. Computer Vision (ICCV ’03), 2003; 1:199-206.
10. Tordoff B, Murray D. Guided Sampling and Consensus for Motion Estimation, Proc. Seventh European Conf. Computer Vision (ECCV ’02), 2002; 1:82-96.
11. Quian X, Ye C. Senior Member, IEEE, “NCC-RANSAC: A Fast Plane Extraction Method for Navigating a Smart Cane for the Visually Impaired, National Science Foundation (IIS-1017672) and NASA.
12. Fischler MA, Bolles RC. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Communications of the ACM, 1981; 24(6):381-395.
13. Nurunnabi A, Belton D, West G. Diagnostic-Robust Statistical Analysis for Local Surface Fitting in 3D Point Cloud Data, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012; 1-3:269-274.