Journal of Advanced Computer Science and Technology, 1 (4) (2012) 284-290

www.sciencepubco.com/index.php/JACST

INAPSAC: A New Robust Inlier Identification Technique

M.Radha[1] and R.Muthukrishnan[2]

Department of Statistics, Bharathiar University, Coimbatore, Tamilnadu, India
[1]radhamyilsamy@gmail.com
[2]muthukrishnan70@rediffmail.com

Abstract

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. These methods tolerate the presence of data points that do not obey the assumed model such points are typically called “outlier”. Recently, various robust statistical methods have been developed and applied to computer vision tasks. Random Sample Consensus (RANSAC) estimators are one of the widely applied to tackle such problems due to its simple implementation and robustness. There have been a number of recent efforts aimed at increasing the efficiency of the basic RANSAC algorithm. N Adjacent Points Sample Consensus (NAPSAC) is one of the RANSAC method used in computer vision task. In this paper a new algorithm is proposed which is the modified version of NAPSAC with 2-sphere method. The accuracy of the proposed algorithm has been studied through a simulation study along with the existing algorithms in the context of RANSAC techniques.

Keywords: *NAPSAC, RANSAC, Robust Statistics*

1 Introduction

Robust estimators were developed and applied in statistics and computer vision during the past decades; most of them can tolerate outliers up to 50%. In computer vision tasks, it frequently happens that outliers and pseudo-outliers occupy the absolute majority of the data. Therefore, the requirement in these robust estimators that occupy outliers less than 50% of all the data points is far from being satisfied for real tasks in computer vision. A good robust estimator should be able to correctly find the fit when outliers occupy a higher percentage of the data. Also, ideally, the estimator should be able to resist the influence of all types of outliers (such as uniformly distributed outliers, clustered outliers and pseudo-outliers).

Fishler and Bolles (1981) were proposed RANdom Sample Consensus algorithms for model fitting with applications of image analysis and other computer vision tasks. The general algorithm of RANSAC is 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 the parameters is evaluated on the input data. Different cost functions have been proposed, the standard being the number of data points consistent with the model. The process is terminated when the probability of finding a better model becomes lower than a user specified probability $\eta 0$. The $1\text{-}\eta 0$ confidence in the solution holds for all levels of contamination of the input data, that is, for any number of outliers within the input data. The performance of RANSAC algorithms 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.

In this paper a new inlier identification method is proposed based on the NAPSAC developed by Myatt and et al (2002). The next section deals with the introduction about the RANSAC and N Adjacent Points SAmple Consensus (NAPSAC). The algorithm of newly proposed robust RANSAC method called Improved NAPSAC (INAPSAC) along with mathematical calculation is presented in section 3. In Section 4 the performance of the proposed algorithm is studied with the existing RANSAC and NAPSAC methods and it is observed that it gives better results in certain situations. The last section provides certain remarks and recommendations which will form a basis for further studies in the field and solve other problems which may occur in practice.

2 Random Sample Consensus techniques

Anton and et al (2005) uses the following strategy for RANSAC technique. 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. As 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.

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

- Random sampling m elements of the input data $S_k \subset x$
- Estimating hypothesis θ_k from S_k
- Measuring the hypothesis score, $R_k=R(\theta_k)$

After generation and evaluation of M hypothesizes, the one with highest score, $\max_{k=1,N}(R_k)$, is selected as the result of robust estimation. Given the expected fraction of inliers γ in the input data and the total number of samples N, the number of algorithm iterations M necessary to find the true model parameters with desired probability P can be calculated.

The algorithm of NAPSAC focused on the efficiency of the sampling strategy. The idea is that inliers tend to be closer to one another than outliers, and the sampling strategy can be modified to exploit the proximity of inliers. This might bring efficient robust recovery of high dimensional models, where the probability of drawing an uncontaminated sample becomes very low even for data sets with relatively low contamination of outliers. Non-uniform sampling has been shown to provide a theoretical advantage over uniform sampling, but will now be shown experimentally to be just as effective in high noise and higher dimensions. To demonstrate this, a simple enhanced sampling algorithm was created. The following algorithm can be used in place of the uniform point sampling process in any of the robust estimation algorithms.

- Select an initial point X_0 randomly from all points.
- Find the set of points, S_{x0}, lying within a hyper sphere of radius r centered on X_0.
- If the number of points in S_{x0} is less than the minimal set size then fail..
- Select points from S_{x0} uniformly until the minimal set has been selected, inclusive of X_0.

The enhanced sampling algorithm was integrated with the RANSAC consensus set cost function to facilitate experimentation. This combination was named N Adjacent Points SAmple Consensus (NAPSAC).

3 Improved N Adjacent Points Sample Consensus (INAPSAC)

Myatt and et.al (2002) proved that the uniform point sampling may be failed in higher dimensions. So, it may be related to its neglect of the spatial relationship between the inlying data points. Using the distribution of the inlying data within the multi-dimensional space to modify the point sampling may improve hypothesis selection. Such models are already used to determine the quality of estimation. The method proposed here is to use a similar technique to select hypotheses through improved point selection. The proposed a new inlier identification scheme based on proximity in three dimensions spheres is as following. Let $\{X_1, X_2, \ldots, X_N\}$be the set of all data points (inliers and outliers). First assume that the initial data point x_0 has already been selected and that point lies on the manifold. So, therefore marginal density of inliers and outliers at a distance r from X_0 can be calculated. Then, by comparing these marginal densities, it can be determined whether selecting by proximity increases the probability of sampling inliers over uniform random sampling. To examine the worst case scenario, where the points are uniformly distributed on the manifold between limits $-t<x<t$ and $-t<y<t$, such that the probability density function of x and y are $p_i(x) = 1/2t$ and $p_i(y) = 1/2t$. Assume that inlier point coordinates are measured with error that satisfies Gaussian distribution with zero mean and standard deviation σ. This distribution is truncated $-t< z< t$, since $t>\sigma$, the truncation will have negligible effect on subsequent calculations. Thus, the joint probability density of the inliers is

$$p_i(x,y,z) = \begin{cases} \frac{1}{4t^2\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right) & if\ -t < x < t, -t < y < t, -t < z < t \\ 0 & otherwise \end{cases}$$

The outliers are uniformly distributed in a 2-sphere centered on the origin radius t then the Probability density function of outliers is

$$p_o(x,y,z) = \begin{cases} \frac{1}{4\pi t^3} & if \quad \sqrt{x^2+y^2+z^2} < t \\ 0 & otherwise \end{cases}$$

Then the joint p.d.f of inliers and outliers is

$P(x,y,z)=\mu p_i(x,y,z)+(1-\mu)p_o(x,y,z)$

In order to find the conditional probability density of selecting an inlying point from a given point on the manifold as a function of their mutual distance, a coordinate transform from Cartesian to Polar is required

$$P_i(r,\varphi) = Kr\exp\left(-\frac{r^2\sin^2\varphi}{2\sigma^2}\right)$$

where $K = \frac{1}{4t^2\sqrt{2\pi}}$

Then the marginal density of inliers is

$$m_i(r) = \int_o^\pi Kr\exp\left(-\frac{r^2\sin^2\varphi}{2\sigma^2}\right)\delta\varphi$$

However, using the trigonometric identities $\sin^2\varphi = \sin 2(\varphi + \pi)$ and $\sin^2\varphi = 1/2\,(1 - \cos 2\varphi)$ this can be rearranged to

$$m_i(r) = Kr\int_o^{2\pi} \exp\left(-\frac{r^2(\frac{1}{2}(1-\cos 2\varphi)}{2\sigma^2} \right)\delta\varphi$$

$$m_i(r) = Kr\int_o^{\pi} \exp\left(-\frac{r^2(1-\cos 2\varphi)}{4\sigma^2} \right)\delta\varphi$$

Integrating the above equation with respect to φ, we get

$$m_i(r) = 2Kr\exp\left(-\frac{r^2}{4\sigma^2} \right)\left[\pi I_0\left(\frac{r^2}{4\sigma^2} \right) \right]$$

where I_0 is the modified Bessel function of first kind. After simplifying the above equation we get marginal density as

$$m_i(r) = \mu\frac{1}{t^2}\sqrt{\frac{\pi}{2}}\left[\frac{r}{\sigma}\exp\left(-\frac{r^2}{4\sigma^2} \right) I_0\left(\frac{r^2}{4\sigma^2} \right) \right]$$

The polar to Cartesian transformation of outliers is

$P_0(r,\varphi) = \frac{3}{4\pi t^3} r$

The marginal density of outliers is

$$m_o(r) = \int_0^{2\pi} \frac{3}{4\pi t^3} r\delta\varphi$$

Integrating the above equation w.r.t φ

$$m_o(r) = (\frac{3}{2t^3} r)(1-\mu)$$

As per Myatt et at (2002) as r increase, $m_i(r) \rightarrow \frac{1}{t^2}$. Consequently, if r_c denotes a distance at which both marginal densities are equal, then

$$r_c \approx \frac{2t}{3}\left(\frac{\mu}{(1-\mu)} \right).$$

The following INAPSAC algorithm is used in the place of sample selection.

- Select initial data point X_0 from *U* at random.
- Find the set of sample points R_{X_0} lying in a 3 dimensional hyper-sphere (2-sphere) of radius r centered on X_0.
- If the size of R_{X_0} is less than the minimal set size m-1 then fail
- Repeat the above steps until the minimal set has been selected including the point X_0.

This results in a cluster of points being selected from a ball. If the initial point, X_0, lies on the manifold, then the rest of the points sampled adjacently will theoretically have a significantly higher probability of being inliers. If there are not enough points within the hyper sphere to estimate the manifold, then that sample is considered a failure.

4 Simulation Study

This section presents the simulation study results to compare the performance of INAPSAC techniques with RANSAC and NAPSAC methods. The simulation study is carried out for different number of threshold such as 2, 4 and 6 and for various sample 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 table. It is observed from the table, in all the situations the INAPSAC technique gives the better result than the other methods, since the no of inlier points are more than the other method.

Table1. The Estimated Results of INAPSAC with other RANSAC Techniques

T	Methods	n=100		n=500		n=100	
		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
4	TLS	0.175107	95	-0.049626	475	-0.033793	950
	RANSAC	-0.603010	96	0.269810	477	0.189419	959
	NAPSAC	-0.442057	84	0.112394	476	1.063397	959
	INAPSAC	0.314658	97	-0.238840	480	0.819703	961
6	TLS	-0.024324	95	-0.017501	475	0.051102	950
	RANSAC	-0.029559	96	1.252636	483	0.953964	962
	NAPSAC	0.121398	96	0.200737	483	-0.381640	963
	INAPSAC	-0.075238	98	-0.038167	485	-1.446939	966

5 Conclusions

In this paper a new RANSAC method is introduced, which is the modified version of NAPSAC method, through a simulation study and it is observed that it gives better results when compared to existing algorithms in certain situations.

INAPSAC has been used extensively and has proven to give better performance than various other robust estimators. The problem of fitting a model with noisy data is still a major and challenging task within the computer vision communities. So, developing a RANSAC method which can tolerate high percentage of outliers is very helpful for researchers. In order to solve problems this may occur in practice.

References

[1] Anton Konouchine, Victor Gaganov and Vladimir Veznevets, "A new maximum likelihood robust estimator", *Proceedings of the International Conference on Computer Graphics and Vision (GrapiCon)*, 2005.

[2] Chum. O and Matas.J, "Randomized RANSAC with sequential probability ratio test", *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV)*, 2005.

[3] Fischler, M.A. and Bolles, R.C, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Commun. ACM*, 24(6), 1981, pp.381-395.

[4] Goldenshluger, A. and Zeevi, A, "The Hough transform estimator", *The Annals of Statistics*, 32(5), pp.1908-1932.

[5] Huber, P.J, *Robust Statistics*, Wiley New York, 1972.

[6] Myatt D.R, P.H.S Torr, S.J. Nasuto, J.M. Bishop, and R. Craddock, "NAPSAC: High noise, high dimensional robust estimation - it's in the bag". Proceedings *of the 13th British Machine Vision Conference (BMVC)*, 2002.

[7] Peter Meer, Doran Mintz, Azriel Rosenfeld, and Dong Yoon Kim, "Robust regression methods for computer vision: A review", *International Journal of Computer Vision*, 6 (1), 1981, pp.59–70.

[8] Rafael C.Gonzalez, Richard E.Woods and Steven L.Eddins, *Digital Image Processing Using MATLAB*, Pearson Education Pte. Ltd, 2004.

[9] Rahul Raguram, Jan-Michael Frahm, and Marc Pollefeys, "A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus", *Proceedings of the European Conference on Computer Vision (ECCV)*, 2008.

[10] Rousseeuw,P.J, "Least median of squares regression", *Journal of the American Statistical Association*, 79(388), 1984, pp.871–880.

[11] Sunglok Choi, Taemin Kim and Wonpil Yu, "Performance Evaluation of RANSAC Family", *Proceedings of the British Machine Vision Conference (BMVC)*, 2009.

[12] Torr, P. and Zisserman, A, "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry", *Computer Vision and Image Understanding*, 78(1), 2000, pp.138-156.