A new inlier identification scheme for robust estimation problems

Wei Zhang and Jana Košecká
Department of Computer Science
George Mason University
Fairfax, Virginia 22030
Email: {wzhang,kosecka}@cs.gmu.edu

Abstract— **Common goal of many computer vision and robotics algorithms is to extract geometric information from the sensory data. Due to noisy measurements and errors in matching or segmentation, the available data are often corrupted with outliers. In such instances robust estimation methods are employed for the problem of parametric model estimation. In the presence of a large fraction of outliers sampling based methods are often the preferred choice. Traditionally used RANSAC algorithm however requires a large number of samples, prior knowledge of the outlier ratio and an additional, difficult to obtain, inlier threshold for hypothesis evaluation.**

To tackle these problems we propose a novel efficient sampling based method for the robust estimation of model parameters. The method is based on the observation that for each data point, the properties of the residual distribution with respect to the generated hypotheses reveal whether the point is an outlier or an inlier. The problem of inlier/outlier identification can then be formulated as a classification problem. The proposed method is demonstrated on motion estimation problems from image correspondences with a large percentage of outliers (70%) on both synthetic and real data and estimation of planar models from range data. The method is shown to be of an order of magnitude more efficient than currently existing methods and does not require a prior knowledge of the outlier ratio and the inlier threshold.

I. INTRODUCTION

Many computer vision and robotics algorithms strive to extract geometric information from the sensory data. This includes images, laser range data, or ultrasound. The geometric information is typically represented by a parametric model. Frequently considered models are planar surfaces from the segmented range data, camera pose from image correspondences or range scans. In most scenarios the data, in addition to the sensor noise, are corrupted with a significant fraction of outliers, due to either measurement errors, mismatches in correspondences or errors in segmentation. This rules out the application of traditional least squares estimators and requires resorting to robust estimation techniques.

In statistics community the efforts focused on obtaining provably robust estimators characterized by their breakdown point. [1] However the achievable breakdown points are usually low and the techniques are very costly to implement in practice. Estimators such as LMedS and LTS [10] can tolerate

only 50% of outliers. Although it is desirable to design estimators with a solid theoretical footing and provable breakdown points, they often have a small bearing on practical problems, which can be tackled reliably. Many of the practical problems have been successfully approached either by sampling based methods (e.g. RANSAC) or Hough Transform, which can empirically tolerate high fractions of outliers.

Our work is motivated by the class of sampling based methods, similar in the spirit to RANSAC algorithm [2] introduced by Fishler and Bolles [2]. When the fraction of outliers is significant and the parametric model is complex, standard RANSAC algorithm requires a large number of samples and an additional, difficult to obtain, inlier threshold for hypothesis evaluation. In the basic algorithm, individual hypotheses generated by the sampling process are evaluated with respect to all data points and ranked based on the number of their inliers, searching for the best hypothesis. The number of needed samples is related to the fraction of outliers which is often not known a-priori. Although RANSAC can handle more then 50% of outliers, as the fraction of outliers increases it becomes prohibitively expensive.

The main contribution of this paper is a novel inlier identification scheme, where we propose to classify the data points directly based on the generated hypotheses. The proposed approach is very efficient, especially for data sets contaminated with large fractions of outliers and eliminates the need for a predefined inlier threshold (scale) and the prior knowledge of the outlier ratio which determines the number of needed samples.

In our work we are motivated and focus on the camera motion estimation problem from image correspondences between two widely separated views. This problem is of great relevance for vision based localization in large scale environments. In particular in the context of the relative positioning or loop closing tasks, the camera pose with respect to a known landmark has to be computed. As Figure 8 demonstrates, in large scale urban environments, the straightforward feature matching stage usually yields a large number of incorrect correspondences. The need for robust estimation methods has been previously explored in this context by several authors [11], [9], [5], [7] and will be reviewed below.

---

[1] The breakdown point of an estimator corresponds to a smallest percentage of outliers, which can cause arbitrarily large values of the estimator.

[2] RANdom SAmple Consensus.

The rest of the paper is organized as follows. In Section 2 we briefly review the basic RANSAC algorithm and discuss its drawbacks. Related work and partial improvements over traditional RANSAC are discussed in Section 3. The proposed hypothesis evaluation and inlier/outlier identification scheme is described in Section 4 and demonstrated on synthetic data. In Section 5 we present experiments on real data and Section 6 concludes the paper.

II. RANSAC ALGORITHM

The essence of the RANSAC algorithm is the generation of multiple hypotheses by means of sampling of the data. Given the minimal number of data points p needed to estimate a parametric model and the fraction of outliers ϵ, we can compute the probability ρ that given m samples, at least one of the samples is outlier free:

$$\rho = 1 - (1 - (1 - \epsilon)^p)^m. \tag{1}$$

In order to achieve a desired probability (confidence) ρ of an outlier free hypothesis and provided that the outlier fraction ϵ is known, one can compute from the above equation the required number of samples:

$$M = \left\lceil \frac{ln(1 - \rho)}{ln(1 - (1 - \epsilon)^p)} \right\rceil. \tag{2}$$

Given the determined number of samples M (calculated based on Equation 2), hypothesis model parameters are estimated for each sample, followed by finding the support (e.g. the number of inliers) for each hypothesis. Alternatively, a stopping criterion can be used to terminate the sampling if sufficient percentage of inliers has been encountered. It has been shown in [5] that the stopping times for the two strategies mentioned above differ only by a multiplicative factor. In the second stage the hypothesis with the largest support is chosen, and all its inliers are used to refine the model parameters. More detailed description of the RANSAC algorithm can be found in [14]. Alternative approach to hypothesis evaluation using spectral graph partitioning techniques in the space of hypotheses generated by sampling has been proposed recently by [8]. Authors' evaluation states that the number of hypotheses needed is on the order of $O(M^2)$, making it exponential in the number of points needed to estimate the model (2 points per line), where M is the number of considered hypotheses. As authors point out, the complexity is similar to RANSAC and clustering is done in the hypothesis space. By studying the distribution of residuals, our method is much more efficient

The larger the sample size p, it is less likely that the sample is outlier free and more samples are needed to achieve the target confidence. For illustration we show the number of samples needed to estimate the fundamental matrix model for displacement between two views. The fundamental matrix has 9 elements, but only 7 degrees of freedom.

When the data set contains 50% of outliers, in order to estimate the fundamental matrix using the linear 8-point algorithm, 766 samples are needed to assure 95% confidence that one outlier free sample is obtained. The number of

Outlier percentage ϵ	30%	40%	50%	60%	70%
7-point algorithm	35	106	382	1827	13696
8-point algorithm	51	177	766	4570	45658

TABLE I

THE THEORETICAL NUMBER OF SAMPLES REQUIRED TO ENSURE 95% CONFIDENCE THAT AT LEAST ONE OUTLIER FREE SAMPLE IS OBTAINED.

required samples goes to 1177 for 99% confidence. As pointed out by [12], the theoretical number of samples is wildly optimistic. In practice, the number of samples required to reach a good hypothesis is around an order of magnitude more. The experiments in [5] also validated this rule. The actual number of samples needed for 99% confidence is on the order of 5000 (our simulations confirm this), which means around 5000 hypotheses need to be evaluated. As shown in Table I, when $\epsilon = 0.7$, the number of required samples is 45658. Consequently, the number of hypotheses to be evaluated will be on the order of 10^5. For each hypothesis, standard RANSAC algorithm computes the residual for every data point. Hence the computation increases linearly with the number of data points. Most of the related work tries to alleviate the efficiency problems related to a large number of required samples, an expensive hypothesis evaluation stage and the inlier threshold selection in various ways.

III. RELATED WORK

Chum and Matas [5] suggested to improve the efficiency of the standard RANSAC by a pre-evaluation, called $T_{d,d}$ test. It exploits the fact that for an erroneous model, only a small number of data points needs to be evaluated. If d randomly selected points pass the $T_{d,d}$ test the hypothesis is not considered further. This enables the authors to increase the efficiency of the hypothesis evaluation stage, but the number of samples remains still large. In [6], the authors proposed to select the sample sets of adjacent points based on the assumption that inliers will tend to be closer to one another than outliers and therefore increasing the probability of an outlier free hypothesis. Guided sampling by quality of matches was proposed by [12] and increased the chance of sampling 'good' correspondences more often and hence generate good hypotheses. Torr and Zisserman [13] have noticed that the simple evaluation of hypotheses by their inlier count is faulty, since it treats all the inliers equally (error terms for all inliers are constant). Consequently, if the threshold T used on residual errors is not set appropriately, the final model estimate will be poor. They suggested using log likelihood of the solution as the support instead of number of inliers. Nister [7] has demonstrated a preemptive RANSAC scheme which runs in real time. The preemptive score is used to sequentially remove bad hypotheses, until only the best hypothesis is left or time budget is used out. The scheme was tested on synthetic data with 20% outliers. In real experiments the points were tracked between individual frame of the video sequence and contained small fraction of outliers. Additional speed up was obtained by the use of the 5-point algorithm

assuming that the camera is calibrated in advance. The issue of the threshold selection for the inlier identification has been addressed recently by [15]. They proposed an automatic scale selection method for estimation of the scale of inliers' noise by analyzing the distribution of residuals of each hypothesis and hence avoiding heuristic threshold selection used for inlier classification. The inlier scale was estimated, using an iterative mean shift algorithm for locating the modes in the residual distribution. Although the approach was capable of handling a large percentage of outliers ($\approx 85\%$) on a simple line fitting examples, the efficiency related to the required number of samples and additional overhead caused by iterative scale estimation have not been addressed.

IV. THE PROPOSED SCHEME

We are motivated by the motion estimation problem from two widely separated views given image correspondences. In this problem the model to be estimated is complex and the data often contain a significant fraction of outliers. The presence of the outliers is particularly pervasive in large scale outdoor urban environments and it is due to a significant viewpoint change, illumination changes and ambiguities due to repetitive structures inherent to buildings. The set of correspondences often contain more than 50% outliers. As the Table 1 indicates using the traditional RANSAC sampling techniques would be prohibitively time consuming, in addition to the issues of inlier threshold selection. Even though the automated threshold selection method [15] can overcome some of the difficulties, it introduces an additional overhead without reducing the number of needed hypotheses.

Note that identification of inliers is at the core of RANSAC algorithm. The final model parameters are then estimated based on the identified inliers. Most of the sampling based algorithms generate many hypotheses which guarantees that with some confidence an outlier free hypothesis is encountered in the set. As shown in Table I, this depends on the complexity of the model and the outlier ratio, which is not known ahead of time. The preemptive RANSAC [7] is the only exception which uses a fixed number of samples (500-800), assuming the outlier percentage is around 20% in a calibrated setting with 5-point algorithm. The idea of the search for good hypotheses remains unchanged. Although this method has been show to work well with video sequence (and hence lower outlier ratios), it has not been extended to data containing more outliers.

In the presented approach, instead of evaluating goodness of individual hypotheses generated by the sampling process, we evaluate the residuals of each data point with respect to all hypotheses. The proposed method is based on the observation, that for each data point the properties (higher order statistics) of the distribution of residuals with respect to generated hypotheses reveal whether the point is an outlier or an inlier. The problem of inlier/outlier identification can then be formulated as a classification problem. The presented approach exploits the fact that even the hypotheses with a small number of outliers contribute to the characteristic

kurtotic shape of inlier's distribution. Hence the presented method does not require per se a presence of an outlier free hypothesis and a large number of samples is not necessary. The approach in addition to its efficiency does not require prior knowledge of the outliers percentage and doesn't need any threshold for identification of inlier's support of the hypothesis. We demonstrate the performance of the proposed method on the problem of motion estimation, with varying outlier percentages (up to 70%) and show that we can correctly identify the inliers over varying fractions of outliers with fixed number of samples. In the next section we will describe the approach and justify it on a simple example. Extensive simulations and experiments on real images are presented in Section 4.

A. Inlier identification procedure

We will describe the proposed method on an example of estimation of the epipolar geometry between two views. Given a set of correspondences $\{\mathbf{x}_i, \mathbf{x}'_i\}_{i=1}^{C}$ between two views of the same scene, our goal is to estimate the fundamental matrix F. Similarly as in the standard RANSAC algorithm we first use sampling to generate a set of hypotheses, in this case fundamental matrices. This is achieved by sampling the set of correspondences by selecting 8-point samples and estimating F using the 8-point algorithm with normalization. At this stage our method dramatically departs from the previously proposed approaches. Instead of evaluating/scoring each hypothesis, we look at the data points directly. For each data point (e.g. correspondence) we study the distribution of residuals with respect to all hypotheses. For a hypothesis F_j instead of considering algebraic residual error $(r_j^i)^2 = (\mathbf{x}_i^T F_j \mathbf{x}'_i)^2$ we use the so called Sampson distance which approximates the reprojection error [3] and is defined as:

$$(r_j^i)^2 = \frac{(\mathbf{x}_i^T F_j \mathbf{x}'_i)^2}{(F_j\mathbf{x}_i)_1^2 + (F_j\mathbf{x}_i)_2^2 + (F_j^T\mathbf{x}'_i)_1^2 + (F_j^t\mathbf{x}'_i)_2^2} \quad (3)$$

where $(F\mathbf{x})_k^2$ represents the square of the k-th entry of the vector $F\mathbf{x}$. Figure 1(a) and Figure 1(b) show the typical error distributions with respect to all generated hypotheses for a data containing 20% outliers. The data was generated using a total of 200 3D points in general position with depth variation of 1000 and projected into two views related by general motion. The inliers were corrupted by a zero mean Gaussian noise and standard deviation of 2 pixels, while the outlier noise was assumed to be uniformly distributed in the interval $[-50, -20] \cup [20, 50]$ pixels. Note that the residual histograms of the inliers and outliers are very different. The inliers typically have strong peaks close to 0, while the outliers don't. We will use this observation for classification of the points to inliers and outliers based on n^{th} order statistics of their residual distributions. Outliers residual histograms can also have high count in the first bin, because some hypotheses are generated using the samples which contain the outlier itself. For this reason the 1st bin was set to 0 prior to the computation of the statistics. The strong peak of inliers error distributions comes from two sources: a particular inlier can be

included in several samples and it can be expected that several good hypotheses yielding a low residual error are included in the hypotheses set. In this example, with 20% of outliers, the probability that a sample of 8 pionts is outlier free is $0.8^8 \approx 0.168$. With 500 generated hypotheses, the expected number of outlier free samples is then $0.168 \times 500 = 84^3$.The number of samples used to generate the hypotheses is set to be $N = 500$. We examine this choice in more detail at the end of this section. Considering the size of the image plane is 400×600, the error histogram has 150 bins, representing the Sampson error ranging from 0 to 149 (large enough to capture the detail of the error distribution). We disregard errors greater than 149.
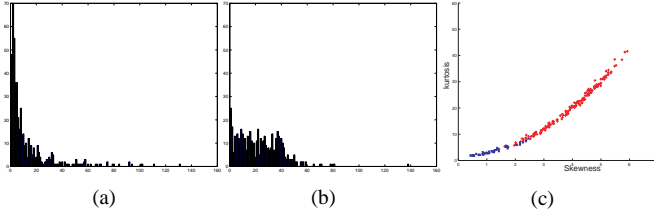

| (a) | (b) | (c) |

Fig. 1. Residual histogram for a true inlier (a) and a true outlier (b), $\epsilon = 0.2$, with each bin corresponds to error in pixels (c): plot of skewness vs. kurtosis computed for all residual distributions of the 200 data points (red '+' represents inliers, while blue 'x' represents outliers.)

1) Features characterizing the distributions: In order to characterize the qualitative differences between the distributions of inliers and outliers depicted in Figure 1, several order statistics can be used. Most commonly used are the lower order statistics such as mean, standard deviation, skewness and kurtosis. Our experiments show that the skewness and kurtosis are very discriminative for the two kinds of residual histograms. Skewness γ measures the asymmetry of the data around the sample mean μ and is defined as

$$\gamma = \frac{E(x-\mu)^3}{\sigma^3}. \tag{4}$$

Skewness of the normal distribution (or any perfectly symmetric distribution) is zero. If the value of skewness is positive, the data are spread out more to the right of the mean than to the left. Kurtosis β is the degree of peakedness of a distribution, which in our case measures how outlier prone a distribution is. Kurtosis is defined as:

$$\beta = \frac{E(x-\mu)^4}{\sigma^4}. \tag{5}$$

For the two histograms shown in Figure 1(a),(b), the kurtosis and skewness for the inlier histogram are 24.4 and 4.6, while for the outlier they are much smaller: 7.6 and 1.7 respectively. These characteristics capture the fact that the inlier's histogram of residuals has much stronger peak than that of an outlier and can be used as feature for further classification.

---

[3]The number of outlier free samples obeys a binomial distribution with N trials and the probability of success is the probability that a sample is outlier free.

We can plot the values of skewness and kurtosis for each data point in 2D, as Figure 1(c) shows. Note that the kurtosis and skewness are correlated, thus it's not necessary to use the two statistics together. In our case only the kurtosis is used for identifying the inliers, making the classification more efficient. From the plot, we can see that the inliers and outliers have different values of skewness and kurtosis. Hence they can be easily separated, either by k-means clustering algorithm or we can simply rank the points in the order of decreasing kurtosis and consider the top k points to be inliers. Notice that the true inliers have kurtosis with much larger variance than the true outliers. Consequently, some true inliers will be misclassified as outliers after the grouping. This however will not cause a problem for the model estimation, because enough true inliers are identified. In case a small number of true outliers is included in the identified inliers set, RANSAC algorithm can be applied for this inliers set. The computational demands are very low, since the outlier percentage is small in this case with no more than 10% outliers as our experiments show. In our experiments we found that using 20 samples was enough to obtain a good hypothesis.

The inlier identification scheme for the case of fundamental matrix estimation is summarized below.

Algorithm 1 Inliers identifications procedure

1) Given the initial set of C correspondences, randomly select N 8-point samples and generate N fundamental matrix hypotheses $\{F_j\}, j = 1, 2 \ldots, N$.
2) For each correspondence (data point), compute its Sampson error $(r_i^j)^2$ with regard to each hypothesis.
3) For each correspondence, estimate its residual distribution by constructing the histogram of N residuals associated with it. The histogram is used to evaluate whether the correspondence is an inlier or an outlier.
4) For C histograms, of residuals compute the value of kurtosis β_k to characterize each of them. In this stage each correspondence is represented by a point in the 1D kurtosis space.
5) Use k-means clustering algorithm to cluster the data into two groups, which are identified inliers and outliers or simply rank the points by their kurtosis value.

Note that the proposed scheme doesn't need a predefined threshold for inliers. The RANSAC schemes require a threshold T to determine whether a data point is an inlier. T is a sensitive parameter and can affect the performance dramatically.

B. Asymptotic running time analysis

Note the steps 3, 4, and 5 of Algorithm 1 require extra computation compared to standard RANSAC. Given N samples and C correspondences, constructing the histograms takes $O(N \times C)$ and computing the value of kurtosis for each takes $O(N \times C)$ multiplications; k-means clustering in one dimension is very efficient. Together, the computation time these steps require is less than the second hypothesis

evaluation stage of standard RANSAC which requires $O(N \times C)$ matrix multiplication. In our experiments the number of samples N was set to be 500. We have also evaluated the sensitivity of our method with respect to the number of samples and obtained repeatable performance for varying outlier ratio when the number of samples varied between 400 to 1000. Note that this is an improvement of an order of magnitude compared to the work reported in [5]. Just for comparison, the standard RANSAC requires $O(N \times C)$ matrix multiplications to evaluate all hypotheses. Without knowing outlier percentage a-priori, number of samples N has to be set conservatively, e.g. $M = 30000$ to handle 60% outliers [16]. Hence the presented approach is more efficient than standard RANSAC, especially when the outlier ratio is high.

C. Justification based on synthetic data

We have shown in Section IV-A a conceptual example that the inliers can be identified directly. In the following section we will demonstrate the feasibility of our approach based on synthetic experiments. 200 correspondences were generated by projecting random cloud of 200 3D points, placed 1000 units of focal length in front of the camera, with the depth variation of 2000. The two views were related by general motion of translation around x-axis and rotation around y-axis of the camera frame. All the correspondences are corrupted by Gaussian noise (standard deviation was 1 pixel). Another 200 random correspondences were uniformly distributed in the image plane, yielding an outlier ratio of $\epsilon = 0.5$. As Figure 2(a) and Figure 2(b) show, the residual distributions for an inlier and outlier are rather different. In this case ($\epsilon = 50\%$). Note that the residual distribution of an inlier is well peaked unimodal distribution, where the mode is close to 0. On the other hand, the distribution of an outlier is more spread out and has multiple modes.
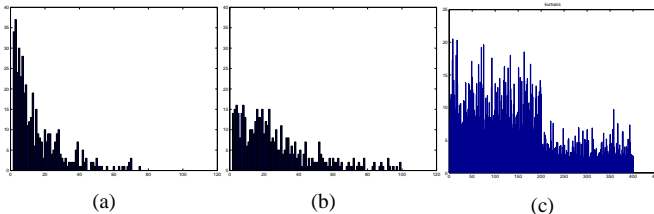

Fig. 2. Residual distribution for a true inlier (a) and a true outlier (b), $\epsilon = 0.5$, (c) kurtosis values of 400 residual distributions one for each point, x-axis corresponds to the point index and y-axis to the corresponding kurtosis value.

Figure 2(c) shows kurtosis of all 400 data points (correspondences). For better visibility, the data are organized as 200 inliers followed by 200 outliers, with x axis being the point index. Note that the inliers and outliers have quite different kurtosis. In this case we have achieved the true positive rate of $138/200 = 68\%$ and false positive rate $2/200 = 1\%$. The inlier identification performs fairly well with this heavily contaminated data set.

As the percentage of outliers increases, it can be expected that the peak of inliers' residual histogram becomes lower and
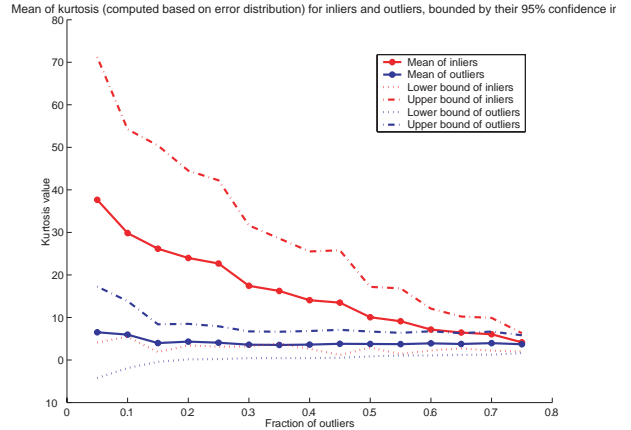

Fig. 3. This figure shows how the kurtosis value changes with different portion of outliers. Mean and 95% confidence interval of inliers' kurtosis are shown in red, mean and 95% confidence interval of outliers' kurtosis are shown in blue.

eventually undistinguishable from the outlier's histogram. It's interesting to see to what extent our approach can tolerate outliers. We tried to study the separability of inliers from the data containing different percentage of outliers. The number of inliers is fixed to be 200 obtained by projecting 200 random 3D points into two widely separated views, while the number of outliers varies to obtain the desired outlier ratio. Figure 3 illustrates the changing kurtosis value. The motion and 3D structure are set the same as in Figure 1. We can see that the kurtosis value of outliers is always small, because they have no significant peaks. The kurtosis of inliers is much larger at first, meaning their error distributions do have strong peaks. Then it decreases as more outliers are added, due to the fact that the outliers are present in a larger number of hypotheses, yielding more dispersed peaks of residuals. When the fraction of outliers ϵ is less than 0.6, the mean of kurtosis computed based on inliers is above 95% confidence interval of that of outliers. Therefore, the kurtosis of a residual histogram associated with inliers and outliers are statistically different, and the inlier group obtained through k-means clustering is very unlikely to contain true outliers. When the outlier percentage increases further but no more than 0.7, the mean of inliers' kurtosis is close to the upper bound of that of outliers'. In this case, the inlier cluster obtained from k-means may contain some true outliers, but the percentage will be much lower than in the original data. As we mentioned before, an additional step of standard RANSAC on the group of identified inliers can yield correct model parameters with a small number of samples. When the outlier percentage grows further to 0.75, inliers and outliers become indistinguishable. Figure depicts the separation of inliers and outliers in the skewness/kurtosis space as the outlier ratio increases. The settings for the experiment were the same as in Figure 1. This indicates that the proposed approach cannot tolerate more than 75% outliers. In theory the standard RANSAC does not have such limitation as long as enough samples are evaluated. When

the outlier ratio ϵ is too high, the required number of samples is so large that it's impractical to work with in practice. As mentioned in Section 2, when $\epsilon = 0.7$, the required number of samples is on the order of a half million, which is too large. The proposed method has a similar working range for the outlier ratio in practice as the standard RANSAC, only that it is much more efficient. Note that the limitation is obtained based on the estimation of fundamental matrix, which requires at least 7 data points. If the model to be estimated is simpler, for instance affine model which requires only 3 data points to estimate, more outliers can be tolerated. The reason is that the required number of samples would decrease dramatically in this case based on the relationship in Equation 2. We will demonstrate this using a simulated example of plane fitting.
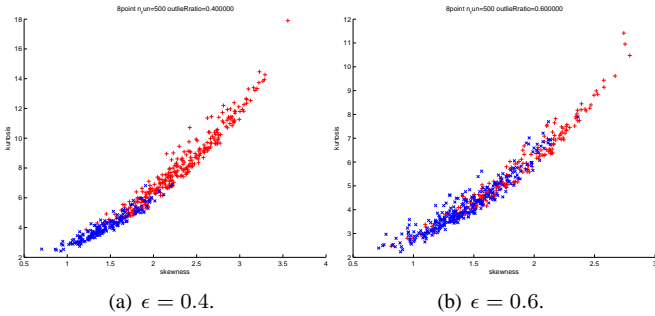

(a) $\epsilon = 0.4$. (b) $\epsilon = 0.6$.

Fig. 4. Plots of kurtosis vs. skewness for different outlier percentage.

1) Plane fitting in 3D space: Figure 5 shows 500 3D data points, where 100 points lie in a plane and are corrupted by Gaussian noise ($\sigma = 1$) and the remaining 400 outliers are uniformly distributed in the space. 108 points are identified as inliers only 10% of which are false positives.
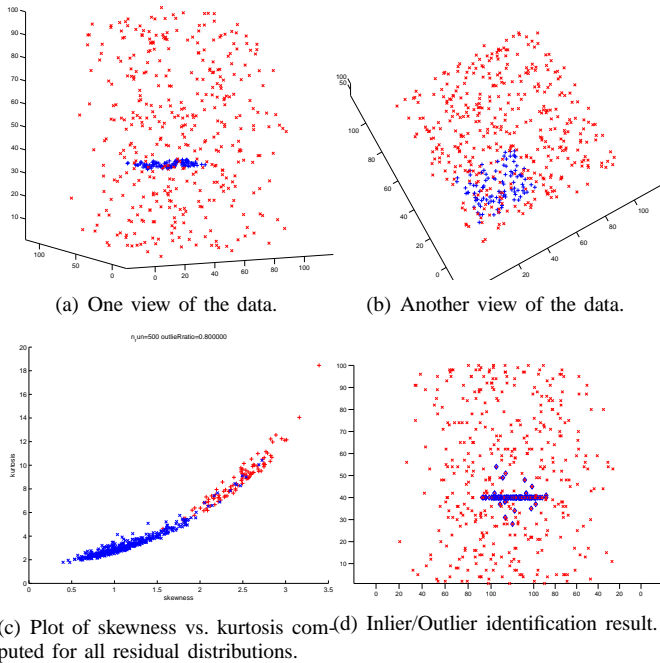

(a) One view of the data. (b) Another view of the data.

(c) Plot of skewness vs. kurtosis computed for all residual distributions. (d) Inlier/Outlier identification result.

Fig. 5. Fitting 3D plane with 80% outliers.

2) Number of samples: Since the method uses all the hypotheses, it is difficult to establish a bound on the number of samples related to a desired confidence as done in standard RANSAC. We study this relationship in simulation by measuring the separability of the inliers and outliers as a function of the number of samples. The Jeffries-Matusita distance measure of separability is used to assess how well two classes may be separated. Assuming that the two classes can be represented by normal distributions with $N(\mu_i, C_i)$ and $N(\mu_j, C_j)$ the Jeffries-Matusita distance is computed as

$$JM_{ij} = \sqrt{2(1 - e^{-\alpha})} \qquad (6)$$

where

$$\alpha = \frac{1}{8}(\mu_i - \mu_j)^T \left(\frac{2}{C_i + C_j}\right)(\mu_i - \mu_j) + \frac{1}{2}ln\left(\frac{\frac{1}{2}|C_i + C_j|}{\sqrt{|C_i||C_j|}}\right)$$

The Jeffries-Matusita distance has an upper bound of $1.414(\sqrt{2})$, and a lower bound of 0. When the calculated distance has a value of the upper bound, the signatures can be said to be totally separable, while the signatures are inseparable when the calculated distance is 0. We experimented to see how the number of samples affects the performance characterized using JM distance. If outliers and inliers are totally separable, their JM distance would be 1.414. On the other hand, if they are tangled together and inseparable, the distance would be 0. For a given outlier ratio, we ran 100 trials of the proposed method and estimated the two clusters based on their kurtosis measure. Mean and standard deviation for both outliers and inliers were obtained in each run. Then we used the average of those means and standard deviations to compute JM distance for the particular number of samples. Figure 6 shows the simulation for data with different outlier ratios. As shown, the separability improves when the number of samples increases. However, after 500 samples, the separability wouldn't improve much as the number of samples increases.

In the case of lower outlier ratios smaller number of samples is sufficient. However lowering the number of samples further will affect the quality of the skewness and kurtosis estimates. Since we would like to have a method which works for a range of outlier ratios we choose the number of samples which has been shown in simulation (Figure 6) to provide a good separation of outliers and inliers for high outlier ratios.

Additional simulations and experiments examining the effect of the number of outliers in each sample as well as reliability of the method depending on the separation of two residual distributions can be found in [17]. The presented approach exploits the fact that even the hypotheses with a small number of outliers contribute to the characteristic kurtotic shape of inlier's distribution.

V. EXPERIMENTS WITH REAL DATA

The proposed scheme was tested with real correspondences sets obtained from a wide baseline matching experiments. The putative correspondences were initiated by matching of SIFT keypoints [4]. Two keypoints are matched when the distance between the two SIFT keypoint descriptor is less
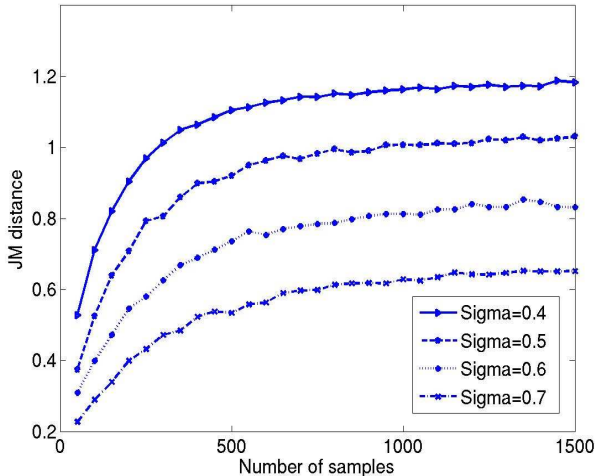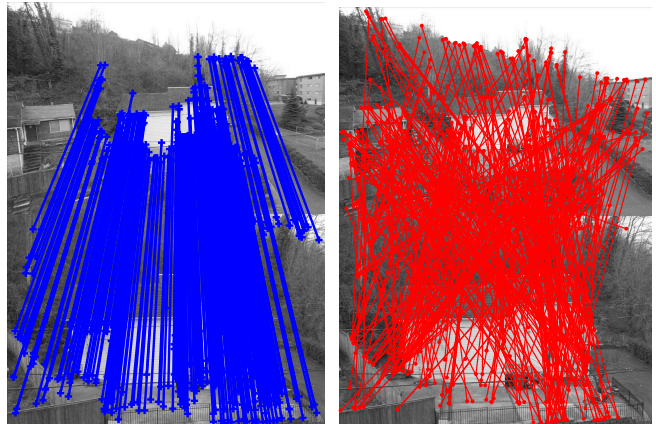
Fig. 6. The JM distance changes with the number of samples. Note that it increases little after 500 samples. The graphs are plotted for different outlier ration Sigma = 0.4,0.5,0.6 and 0.7.

(a) identified inliers.　　　　(b) identified outliers.

Fig. 7. 750 correspondences are initiated with around 50% outliers. 364 inliers are identified without false positive.

(a) identified inliers.　　　　(b) identified outliers.

Fig. 8. 383 correspondences are initiated with approximately 60% outliers. 93 inliers are identified with only 1 false positive. Note the first left door in the left image corresponds to second left door in the right image.

than some threshold τ. We ran extensive experiments with correspondence sets containing different portion of outliers. We tested the method in urban environment using a wide baseline matching between two views In addition to a large change of viewpoint between the views, these scenes contain many repetitive structures, making the problem of finding correspondences by means of matching local feature descriptors highly ambiguous. Our focus is on the inlier identification capability of the proposed scheme. The identified inliers are not refined with additional RANSAC, so they might still contain few true outliers for severely contaminated data sets.

When the percentage of outliers is low, our approach can identify inliers and outliers directly almost without a mistake. The low percentage of outliers can also be handled by RANSAC without an excessive computational overhead. We emphasize that our approach is suitable for correspondence sets with a significant portion of outliers of more than 40%. Three examples are shown in Figures 8, 7 and 9. The identified inlier sets include most of the true inliers with a very few outliers.

It is known that the distance threshold τ used for matching the SIFT keypoint descriptors affects the number of matches. Loose threshold results in many false correspondences. If the threshold is set too tight, hardly any matches could be found. Our work suggest a straightforward way to handle this: set a relatively loose threshold to obtain initial set of correspondences and apply the proposed scheme to identify the true inliers. The presented approach was evaluated on the data set of outdoors street level images used in ICCV Vision Contest provided at (http://research.microsoft.com/iccv2005/Contest/). The goal of the contest was to compute GPS locations of a set of unlabeled images, given a set of views tagged with GPS coordinates. This required, given a new unlabeled query view to compute the pose of the camera with respect to the two
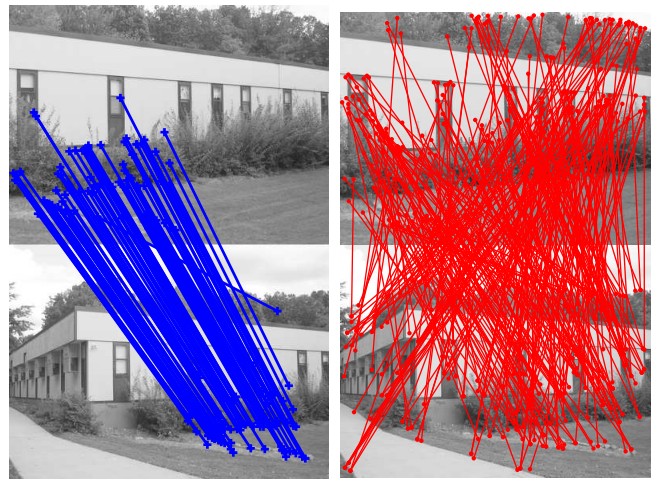
nearest labeled reference views and triangulate the obtained position. The presented approach was used in a winning entry in the contest and reported superior performance and efficiency compared to other methods. The absolute computation time as well as accuracy of the localization method compared to the ground truth can be found at [1].

VI. Conclusion and future work

In this paper we proposed a new inlier identification scheme for robust estimation problems. We have demonstrated that it can efficiently handle data sets containing significant level of outliers. Inliers can be identified directly without looking for good hypothesis, thus avoiding the need for large number of samples, which is required for the standard RANSAC algorithm. In addition to the efficiency of the proposed approach, we have also eliminated the need for sensitive threshold selection for the outlier identification as well as the need for prior
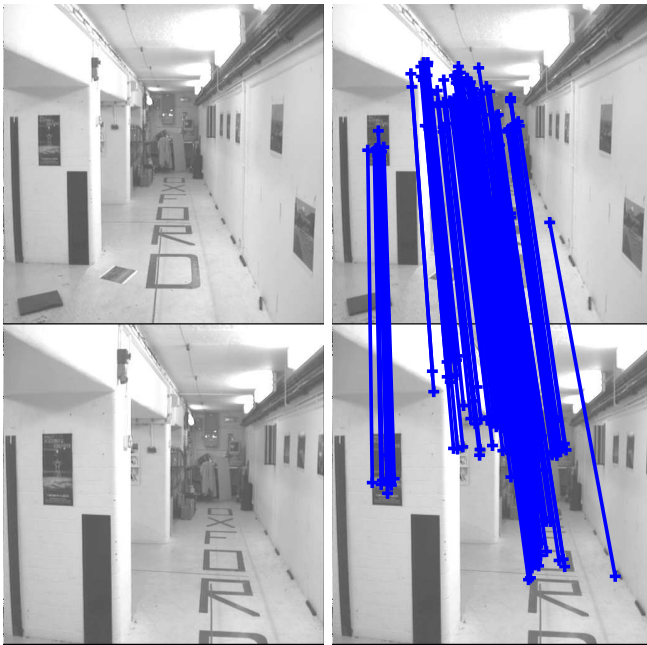
| (a) Pair of images. | (b) identified inliers. |

Fig. 9. Two frames of the widely used Corridor sequence (bt.001 and bt.006), obtained from http://www.robots.ox.ac.uk/ vgg/data/. Outlier percentage is over 50%. 134 inliers are identified with no false positive.

knowledge about the percentage of outliers. We would like to emphasize that the proposed method is especially suitable for the data with a large number of outliers as demonstrated in our application and often encountered in wide baseline matching. The proposed scheme was tested extensively with both synthetic and real data. We plan to refine the inlier identification step in the future, by replacing the k-means clustering by its probabilistic version and hence obtaining the probability of being an inlier for each data point. We are also in the process of carrying out more extensive experiments with different distributions of outliers, in order to asses the generality of the presented method.

REFERENCES

[1] ICCV 2005. Computer Vision Contest. In *http://research.microsoft.com/iccv2005/Contest/Results/Results5Final.htm*, 2005.
[2] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In *Proceedings of the 4th European Conference on Computer Vision*, pages 683 – 695, 1996.
[3] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, 2000.
[4] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004.
[5] Jiri Matas and Ondrej Chum. Randomized ransac with T(d,d) test. In *BMVC'02*, pages 448–457, 2002.
[6] D. Myatt, P. Torr, S. Nasuto, and R. Craddock. Napsac: High noise, high dimensional model parameterisation - its in the bag. In *BMVC*, pages 458–467, 2002.
[7] David Nister. Preemptive ransac for live structure and motion estimation. In *ICCV'03*, pages 199–206, 2003.
[8] E. Olson, M. Walter, S. Teller, and J. Leonard. Single cluster spectral graph partitioning for robotics applications. In *Robotics Science and Systems*, 2005.
[9] Philip Pritchett and Andrew Zisserman. Wide baseline stereo matching. In *ICCV'98*.
[10] Peter J. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Association.*, 79:871–880, 1984.
[11] F. Schaffalitzky and A. Zisserman. Viewpoint invariant texture matching and wide baseline stereo. In *ICCV*, July 2001.
[12] Ben Tordoff and David W. Murray. Guided sampling and consensus for motion estimation. In *ECCV (1)*, pages 82–98, 2002.
[13] P. Torr and A. Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *CVIU*, 78:138–156, 2000.
[14] Philip H. S. Torr, Andrew Zisserman, and Stephen J. Maybank. Robust detection of degenerate configurations whilst estimating the fundamental matrix. pages 312–333, 1998.
[15] H. Wang and D. Suter. Robust adaptive-scale parametric model estimation for computer vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(11):1459–1474, 2004.
[16] H. Wang and D. Suter. Robust fitting by adaptive-scale residual consensus. In *ECCV*, pages 107–118, 2004.
[17] W. Zhang and J. Košecká. Ensemble method for robust motion estimation. In *Workshop on 25 years of RANSAC, IEEE Conference on Computer Vision and Pattern Recognition*, 2006.