Department of Computer Science
George Mason University
Technical Report Series

4400 University Drive MS#4A5
Fairfax, VA 22030-4444 USA
http://cs.gmu.edu/
703-993-1530

A new inlier identification scheme for robust estimation problems

Wei Zhang
wzhang2@cs.gmu.edu

Jana Kosecka
kosecka@cs.gmu.edu

Technical Report GMU-CS-TR-2005-7

Abstract

The data in vision problems, often heavily contaminated by outliers, call for efficient robust techniques to identify inliers for correct estimation. RANSAC algorithm is a frequently used robust estimator for computer vision problems. In traditional RANSAC scheme, when data contain significant fraction of outliers, large number of samples is needed in order to obtain at least one outlier free sample. In addition to that, each hypothesis generated from the samples is typically evaluated using all data points, which further lowers the efficiency. In this paper, we propose a novel hypothesis evaluation scheme, which enables efficient classification of the data points as outliers and inliers, while requiring a small fixed number of samples. The method is based on the observation that for each data point the properties of the distribution of the residuals with respect to the generated hypotheses reveal whether the point is an outlier or inlier. The problem of inlier/outlier identification can then be formulated as a classification problem. We demonstrate the proposed method on motion estimation problems with large fraction of outliers on both synthetic and real data.

1 Introduction

Many computer vision tasks need to deal with data set contaminated by outliers, thus entail a robust estimation step. RANSAC[1] algorithm introduced by Fishler and Bolles [1] has been widely used for for various robust estimation problems in computer vision community. In our work we are motivated and focus on the problem of estimation of camera motion from correspondence between two widely separated views, which usually contains large number of incorrect correspondences. The need for robust estimation methods has been previously explored in this context by several authors [2, 3, 4].

The contribution of our paper is a novel inlier identification scheme for dealing with heavily contaminated data, so that correct estimation can proceed based on it, without interferences from outliers. In the traditional RANSAC algorithm, the hypothesis evaluation stage is essential for identifying which data points are outliers and inliers. The stage is often time consuming, especially for heavily contaminated data. Instead of evaluating individual hypothesis generated by the sampling process, we propose to classify the data points directly based on the generated hypotheses. The proposed approach is very efficient especially for data sets heavily contaminated with outliers, which cause the standard RANSAC to run inefficiently. When the percentage of outliers is low, the existing RANSAC schemes work as efficiently as the proposed method does. In this paper we focus on the motion estimation problem between two widely separated views. The approach is however applicable to any robust estimation problem. In addition to the efficiency advantage, the proposed scheme works without the need of predefined inlier scale. The predefined inlier scale is a sensitive parameter, recently some work [5] has been proposed for automatic inlier scale (threshold) selection.

The rest of the paper is organized as follows. In Section 2 we briefly review the basic RANSAC algorithm and discuss its efficiency. Related work is 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

---

[1] RANdom SAmple Consensus.

present experiments on real data and Section 6 concludes the paper.

2 RANSAC algorithm

The essence of the RANSAC algorithm is the generation of multiple hypotheses by means of sampling of the data. Given a predetermined number of samples M (calculated based on Equation 2), hypothesis model parameters are estimated for each sample, followed by finding the support (typically 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 [6] that the stoping 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 many papers such as [7].

Given m samples, the probability ρ that they include at least one outlier free sample is:

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

where p is the number of points per sample, and ϵ is the outliers percentage. Based on this relation, the required number of samples M can be computed theoretically for giving ϵ and ρ:

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

Typically, p is the minimum number of data points needed to compute the model parameters. The larger the sample size, the less likely that the sample is outlier free. Consequently, more samples are needed to achieve a target confidence. For illustration, in the context of wide baseline matching and motion estimation, fundamental matrix is the model to be estimated. The fundamental matrix has 9 elements, but only 7 degrees of freedom.

When data set contains significant portion of outliers, for example, 50% outliers, to estimate fundamental matrix using linear 8-point algorithm, we need 766 samples to assure 95% confidence that one outlier free sample is obtained. The number of required samples goes to 1177 for 99% confidence. As pointed out by [8], the theoretical number of iterations is wildly optimistic. In practice, the number of iterations required to reach a good hypothesis is around an order of magnitude more. The experiment in [6] also validated this rule. Thus the actual number of samples needed is on the order of 5000 (our simulations confirm this), which means around 5000 hypotheses need to be evaluated. Therefore, RANSAC would cause significant delay, which would rise even more dramatically for data with more outliers. As shown in Table 1, when $\epsilon = 0.7$, the number of required samples is 45658. Consequently, almost a half million hypotheses are needed to be evaluated to obtain a good motion estimation. Even if the 7-point algorithm (which requires less number of samples) is used instead, the actual number of hypotheses which needs to be evaluated is at the same level as that of 8-point algorithm. The reason is that 7-point algorithm returns up to 3 solutions (hypotheses), while 8-point algorithm only returns 1 solution. Besides the problem caused by large portion of outliers, large data sets will also cause additional delay. For each hypothesis, standard RANSAC computes error for every data point. Hence the computation increases linearly with the number of data points. Most of the related work tries to alleviate these problems in various ways.

3 Related work

Chum and Matas [6] suggested to improve the efficiency of the standard RANSAC by a pre-evaluation, called $T_{d,d}$ test. Even though the number of samples increased a lot, time is saved in the hypothesis evaluation scheme, because only a fraction of data points was be evaluated. In [9], the authors proposed to select sample sets of adjacent points based on the assumption that inliers will tend to be closer to one another than outliers. Guided sampling by quality of matches was proposed by [8]. Torr and Zisserman [10] have noticed that the traditional inlier count approach is faulty, since it treats all the inliers equally (error terms for the inliers are constant). Consequently, if the threshold T on the residual errors which is used for classifying the data points as inliers and outliers is not set appropriately, the final model estimate will be poor. They suggested using log likelihood of the solution as support instead of number of inliers. Nister [11] has demonstrated a preemptive RANSAC scheme which can run in real time, preemptive score was 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, and tracking results of a real video sequence. Additional speed up was obtained by the use of the 5-point algorithm method assuming camera is calibrated in advance.

4 The proposed scheme

Our approach is motivated by the wide baseline matching problem. Due to the often significant viewpoint change and additional illumination changes and ambiguities due to the repetitive structures, finding correspondences is often difficult in the absence of some geometric model.

Outlier ratio ϵ	20%	30%	40%	50%	60%	70%
7-point algorithm	13	35	106	382	1827	13696
8-point algorithm	17	51	177	766	4570	45658

Table 1: The theoretical number of samples required to ensure 95% confidence that at least one outlier free sample is obtained. The actual required number of samples is much higher.

The sets of correspondences contain many outliers, often more than 50%. Using the traditional sampling techniques would be very time consuming in order to obtain good outlier free hypothesis.

Identifying inliers is the core of RANSAC algorithm. The final model parameters are then estimated based on those identified inliers. Basically, all the existing versions of RANSAC determine inliers evaluating all the generated hypotheses and looking for a good one. As described in Section 2, large number of samples is needed to obtain a good hypothesis. The preemptive RANSAC [11] is the only exception which uses a fixed number of samples (500-800), assuming outlier fraction around 20% and calibrated setting with 5-point algorithm. The essence of the preemptive RANSAC scheme is still try to find the good hypotheses. Although this method has been show to work well with video sequence, it is not necessarily extendable to data containing more outliers taking with uncalibrated camera.

The efficiency of our approach comes from the fact that we try to *identify inliers directly* bypassing the hypothesis evaluation stage. The proposed method is based on the observation, that for each data point the properties of the distribution of the residuals with respect to the generated hypotheses reveal whether the point is an outlier or inlier. The problem of inlier/outlier identification can then be formulated as a classification problem. Since we do not require good hypothesis to be generated, 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. In the next section we will describe the approach and justify it on a simple example.

4.1 Inlier identification procedure

We will describe the proposed methods on an example of estimation of the epipolar geometry between two views. Given the correspondences set $\{\mathbf{x}_i, \mathbf{x}'_i\}$ between two views of the same scene, let the number of correspondences to be C. Our goal is to estimate the fundamental matrix. Similarly as in the standard RANSAC scheme we first use sampling to generate a set of hypotheses (*i.e.*fundamental matrices). This is achieved by sampling the set of correspondences by selecting 8-points samples and estimating F using 8-point algorithm with normalization. At this stage our method dramatically departs from the other 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 the errors with respect to all hypotheses. For a hypothesis F_j instead of considering 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 geometric distance of the point from the epipolar line and is defined as

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

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)shows 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 projected into two views related by general motion. Note that the error histograms of inliers and outliers are very different. We will use this observation for classification of the points to inliers and outliers based on nth order statistics of their residual distribution. The inliers typically have strong peaks close to 0, while the outliers don't. The outlier can also has high value in the first bin, because some hypotheses are generated using samples which contain the outlier itself. For this reason the 1^{st} bin was set to 0 prior to computation of the statistics. The strong peak of the inlier's error distribution comes from two sources: the 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 the probability that an 8-point sample is outlier free is $0.8^8 \approx 0.168$, the expect number of outlier free samples is approximately $0.168 \times 500 = 84$. [2] The number of samples used to generate the hypotheses is set to be $N = 500$. Based on our experiments, 500 samples are enough to approximate the error histogram, which is the foundation for inlier identification. The error histogram has 150 bins, representing Sampson error ranging from 0 to 149 (large enough to capture the detail of the error distribution). Errors greater than 149 are accumulated into the 150^{th} bin. The 150^{th} bin sometimes has a

---

[2] 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.

large value, which will affect the statistics that we want to measure. Due to this fact it is not considered for computing the statistics.

4.1.1 Features for characterizing distributions

Now we know error histograms of inliers and outliers are different, some features are needed to be extracted to capture them. For distributions several order statistics can be used to capture the qualitative properties of the distribution. Most commonly used are the lower order statistics such as mean, standard deviation, skewness and kurtosis. Our experiments show that skewness and kurtosis are very discriminative for the two kinds of error histograms. Skewness γ measures the asymmetry of the data around the sample mean μ

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

The skewness of the normal distribution (or any perfectly symmetric distribution) is zero. If 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, the kurtosis and skewness for the inlier 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 inlier's error histogram has much stronger peak than that of an outlier and can be used as feature for further classification.

Using skewness and kurtosis as features for each data point (correspondence) we can plot distribution of data in the 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. Only kurtosis is used for identifying the inliers, making the classification more efficient. From the plot, we can see that the two groups of points locate in different range. Therefore, we can use k-means clustering algorithm to divide data into two groups. The group with larger kurtosis are the identified inliers. Notice that the true inliers have kurtosis with much larger variance than that of true outliers. Consequently, some true inliers will be misclassified as outliers after the grouping. But this wont's cause problem for model estimation, because enough true inliers are identified. Also, a small number of true outliers might be included in the identified inliers set. The standard RANSAC can be applied for this inliers set, the computation is trivial, since outlier percentage is very low in this case[3] with no more than 10% outliers as our experiments show.

The inlier identification procedure is summarized below (for the case of fundamental matrix estimation).

Algorithm 1 Inliers identifications procedure

1. 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 [12] r_i^j with regard to every hypothesis.
3. For each correspondence, estimate its error distribution by constructing histogram of the N errors associated with it. The histogram is used to evaluate whether the correspondence is an inlier. A total of C histograms are constructed.
4. For the C error histograms, compute 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 some classication algorithm (k-means clustering algorithm in our experiments) to cluster the data into two groups, which are identified inliers and outliers.

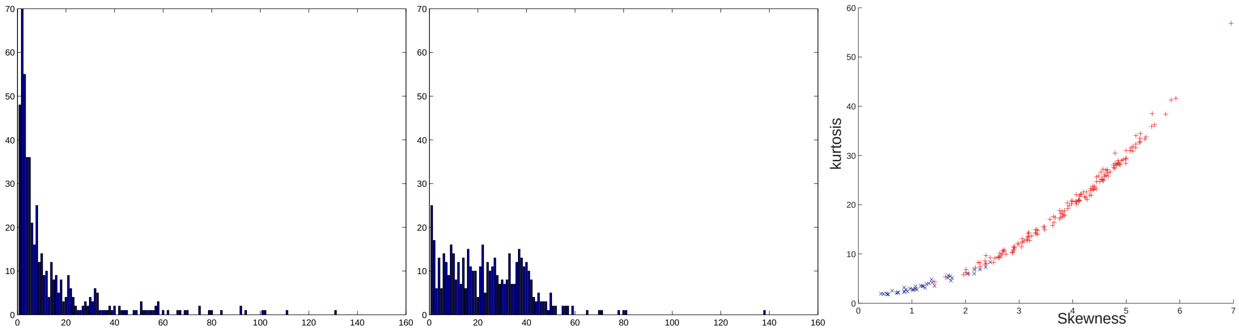
Note that the proposed scheme doesn't need a predefined threshold for inliers. The RANSAC schemes require a threshold T to determine whether a data is inlier. As mentioned in [10], the choice of the threshold is a sensitive parameter and can affect the performance dramatically. Without the need for the predefined T makes the proposed scheme very flexible to handle different data, which shows clear advantage over standard RANSAC scheme.

4.2 Asymptotic running time analysis

Note the 3, 4, and 5 steps require extra computation than standard RANSAC. Constructing histogram takes $O(N \times C)$ additions, and computing kurtosis takes $O(N \times C)$ multiplications. k-means clustering in one dimension is very efficient. Together, the computation time they require are less than the second hypothesis evaluation stage of standard RANSAC which requires $O(N \times C)$ matrix multiplication. N is set to be 500 throughout our experiments. Standard RANSAC requires $O(M \times C)$ matrix multiplication to evaluate all

---

[3] This step is not used in our inlier identification scheme, we emphasize in the inlier identification scheme.



(a) (b) (c)

Figure 1: Error distribution for a true inlier (a) and a true outlier (b), $\epsilon = 0.2$. (c): plot of skewness vs. kurtosis computed for all residual distributions of the 200 data points (red '+' represents inliers, while blue 'x' represents outliers.

the hypotheses. Without knowing outlier percentage a-priori, M has to be set conservatively, e.g. $M = 50000$ to handle 60% outliers. Therefore, M is usually much larger than 500 to deal with real data. *i.e.*, our approach is much more efficient than standard RANSAC, especially when the outlier percentage is high.

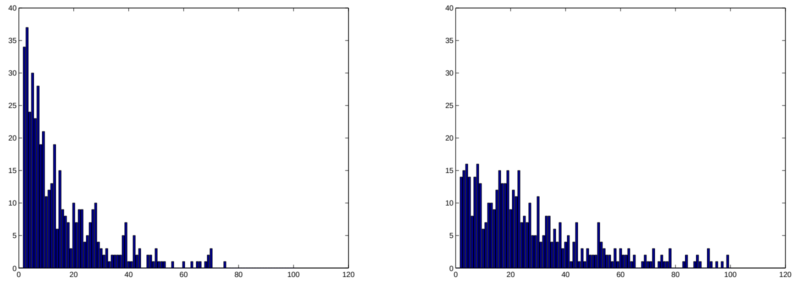
4.3 Justification based on synthetic data

We have shown in Section 4.1 conceptual example that inliers can be identified directly. Now we demonstrate the feasibility of our approach based on a synthetic correspondences set with 200 correspondences (generated by projecting 200 random 3D point into two views), 200 random correspondences are added as outliers (total 400 data with $\epsilon = 0.5$). All the correspondences are corrupted by Gaussian noise (standard deviation was 1 pixel).

As Figure 2(a) and Figure 2(b) show, error distribution for inlier and outliers are rather different in this case ($\epsilon = 50\%$). This can be explained as follow: the residual distribution of each point is a mixture of two distributions. Residuals of wrong hypotheses are proximately random because wrong hypotheses are computed based on random combination of inliers and outliers; while residuals of correct hypotheses are coherent and close together, because correct hypotheses are computed based on inliers only. Thus residual distributions of inliers are well peaked unimodal distribution, where the mode is close to 0 because residuals of inlier to correct hypotheses should be close to 0. On the other hand, for the outliers the distribution are distributed and has multiple modes.

Figure 3 shows kurtosis of all the 400 data points (correspondences). For better visibility, the data are organized as 200 inliers followed by 200 outliers. As is shown, inliers and outliers have quite different kurtosis. Inlier identification by k-means clustering is shown in Figure 4. The true inliers are represented by "x", the 138 identified inliers are circled. 2 false positives (outliers are identified as inliers) are included which are colored red. Other outliers are not shown for better visibility. With 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 peak of inliers' error histogram becomes lower and eventually undistinguishable from the outlier. It's interesting to see to what extent our approach can tolerate outliers. We tried to study the separability of inliers from 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 for desired percentage. Figure 5 illustrates the changing of kurtosis. We can see kurtosis 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, because outliers would disperse the peaks. When fraction of outliers ϵ is less than 0.6, the mean of kurtosis computed based on inliers is above the 95% confidence interval of that of outliers. Therefore, the kurtosis of error histogram associated with inliers and outliers are statistically different, and 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, mean of inliers' kurtosis is close to 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 this inlier group can obtain model parameter with a small num-



(a) (b)

Figure 2: Error distribution for a true inlier (a) and a true outlier (b), $\epsilon = 0.5$.

ber of samples. When outlier percentage goes further to 0.75, inliers and outliers become indistinguishable. This indicate that the proposed approach can not tolerate more than 75% outliers. Standard RANSAC hasn't such limitation theoretically as long as enough samples are evaluated, but when ϵ is too high, the required number of samples is so large that it's impractical to work 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 already too huge to work. So the proposed method has the same working range as standard RANSAC in practice, only that its 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, please refer to Equation 2.

5 Experiments with real data

The proposed scheme was tested with real correspondences sets obtained from wide baseline matching. The putative correspondences were initiated based on matching of SIFT keypoints [13]. Two keypoints are set to be correspondence when the distance between the two SIFT keypoint descriptor is less than some threshold τ. We ran extensive experiments with correspondences sets containing different portion of outliers. We tested the methods in the domain of matching two views of urban scenes and/or buildings. In addition to large change of viewpoint between the views, these scene 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 5 steps scheme. The identified inliers are not refined with additional RANSAC, so they might still contain few true outliers for severely contaminated data sets.

5.1 Handling data sets with large number of correspondences

Additional concern we point out in this context is the issue of handling the data sets with large number of correspondences. As mentioned before, when every data point need to be evaluated, RANSAC will be slower. In our setting we adopt a strategy that sampling can be guided by match score [8] between the two descriptors. This assumes that the score reflects the likelihood of good match. Instead of assigning sampling sampling the entire set of correspondences, we pick a fixed number L of matches with highest matching scores. This is same as setting sampling probability to be 0 for those low quality match, which is a simpler strategy. Though it seems that many data points got wasted, we save time by not sampling the likely outliers. In case the number of matches is high, L shouldn't be small. We use $L = 100$ in our experiments. Even if original data has 60% outliers, 40 inliers can still be expected, which is enough for further process. The actual inliers is usually more, since high score match are likely to be inliers.

When percentage of outliers is low, our approach can identify inliers and outliers directly almost without mistake. This is not very interesting, since standard RANSAC can also handle them very well. We emphasize on correspondences sets with significant portion of outliers, no less than 40%. The results are very promising. Three examples are shown in Figures 8, 6 and 7. The identified inlier sets include most true inliers with very few outliers. Figure 9 shows that instead of using all available data, we can use only part of likely better data. Therefore, the data set with very large number of correspondences won't cause too much delay.

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.

6 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 standard RANSAC algorithm. In addition to the efficiency of the proposed approach, we have also eliminated the need for sensitive threshold selection for outlier identification as well as need for prior knowledge about the percentage of outliers (which is needed when fixed number of samples is used in standard RANSAC). We would like to emphasize that the proposed method is especially suitable for data with large number of outliers as motivated and demonstrated in our application and often encountered in wide baseline matching. The proposed scheme is tested extensively with both synthetic and real data. We plan to refine the inlier identification step in 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] Fischler, M.A., Bolles, R.C.: 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. (1996) 683 – 695

[2] Schaffalitzky, F., Zisserman, A.: Viewpoint invariant texture matching and wide baseline stereo. In: ICCV. (2001)

[3] Pritchett, P., Zisserman, A.: Wide baseline stereo matching. (In: ICCV'98)

[4] Goedeme, T., Tuytelaars, T.: Fast wide baseline matching for visual navigation. In: CVPR'04. (2004) 24 – 29

[5] Wang, H., Suter, D.: Robust adaptive-scale parametric model estimation for computer vision. IEEE Trans. Pattern Anal. Mach. Intell. **26** (2004) 1459–1474

[6] Matas, J., Chum, O.: Randomized ransac with t(d,d) test. (In: BMVC'02)

[7] Torr, P.H.S., Zisserman, A., Maybank, S.J.: Robust detection of degenerate configurations whilst estimating the fundamental matrix. (1998) 312–333

[8] Tordoff, B., Murray, D.W.: Guided sampling and consensus for motion estimation. In: ECCV (1). (2002) 82–98

[9] Myatt, D., Torr, P., Nasuto, S., Craddock, R.: Napsac: High noise, high dimensional model parameterisation - its in the bag. In: BMVC. (2002) 458–467

[10] Torr, P., Zisserman, A.: Mlesac: A new robust estimator with application to estimating image geometry. CVIU **78** (2000) 138–156

[11] Nister, D.: Preemptive ransac for live structure and motion estimation. In: ICCV'03. (2003) 199–206

[12] Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN: 0521540518 (2000)

[13] Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (2004)

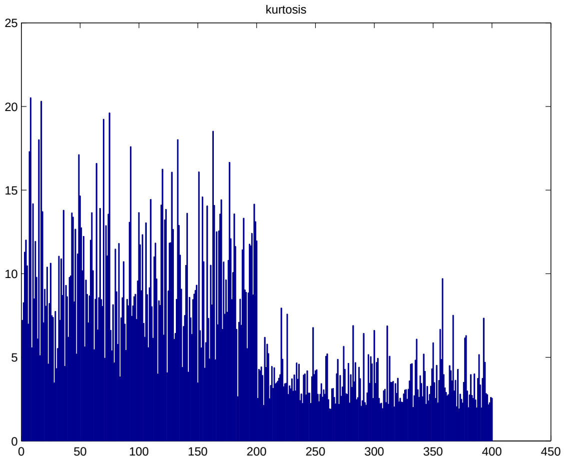


Figure 3: The kurtosis of the 400 error distributions.

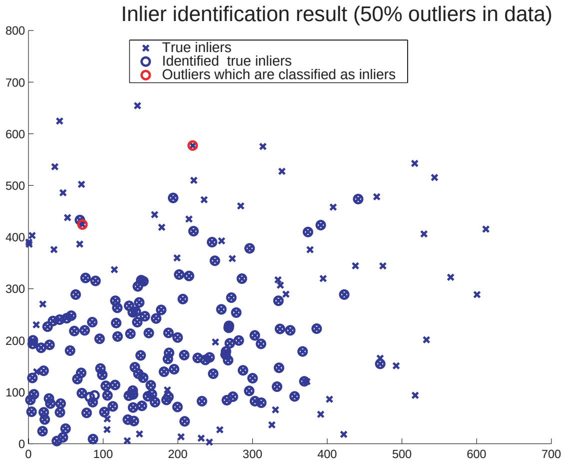


Figure 4: The classification result based on kurtosis.

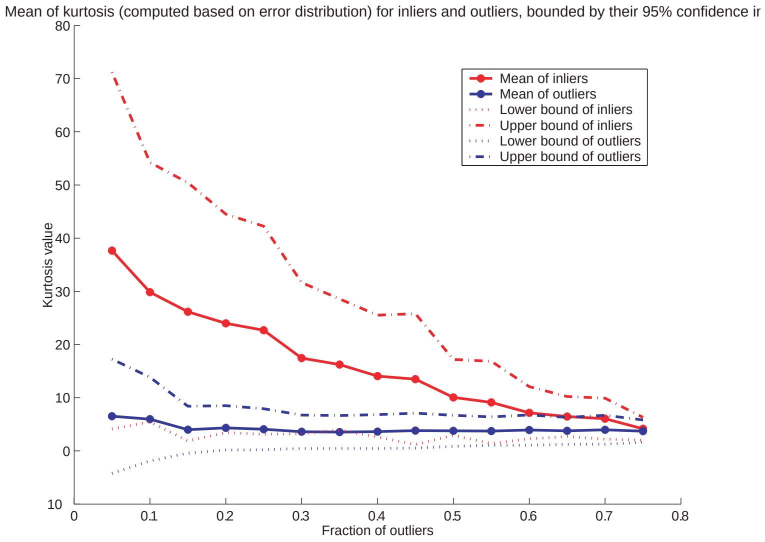
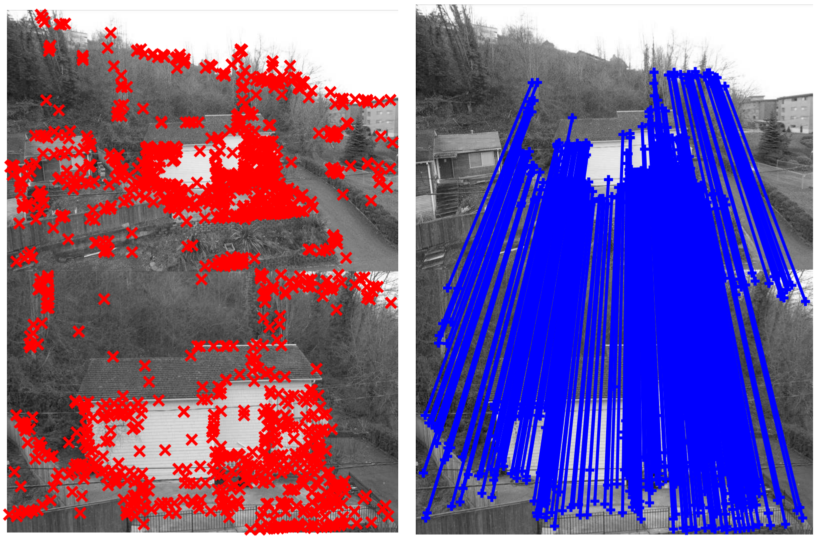
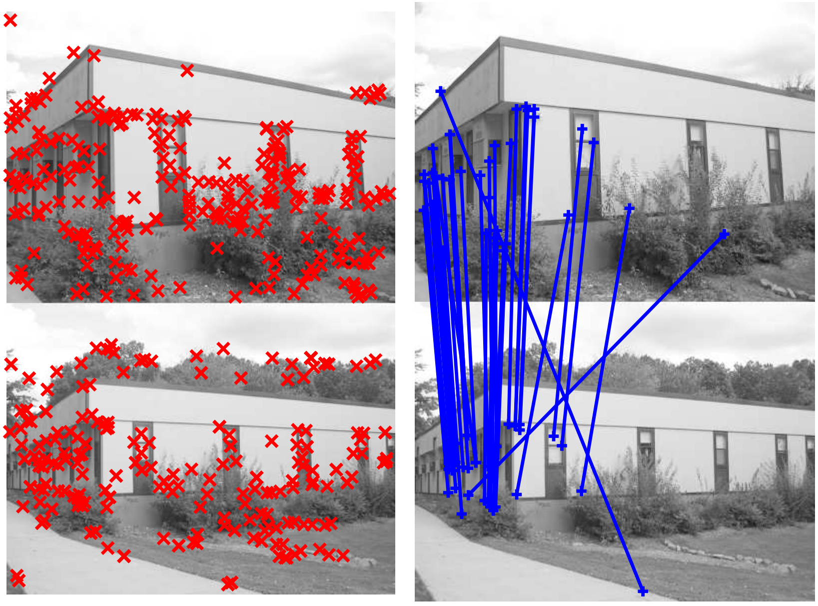


Figure 5: This figure shows how the kurtosis value changes with different portion of outliers. Mean and 95% confidence interval of inliers' kurtosis are shown in red, mean and 95% confidence interval of outliers' kurtosis are shown in blue.



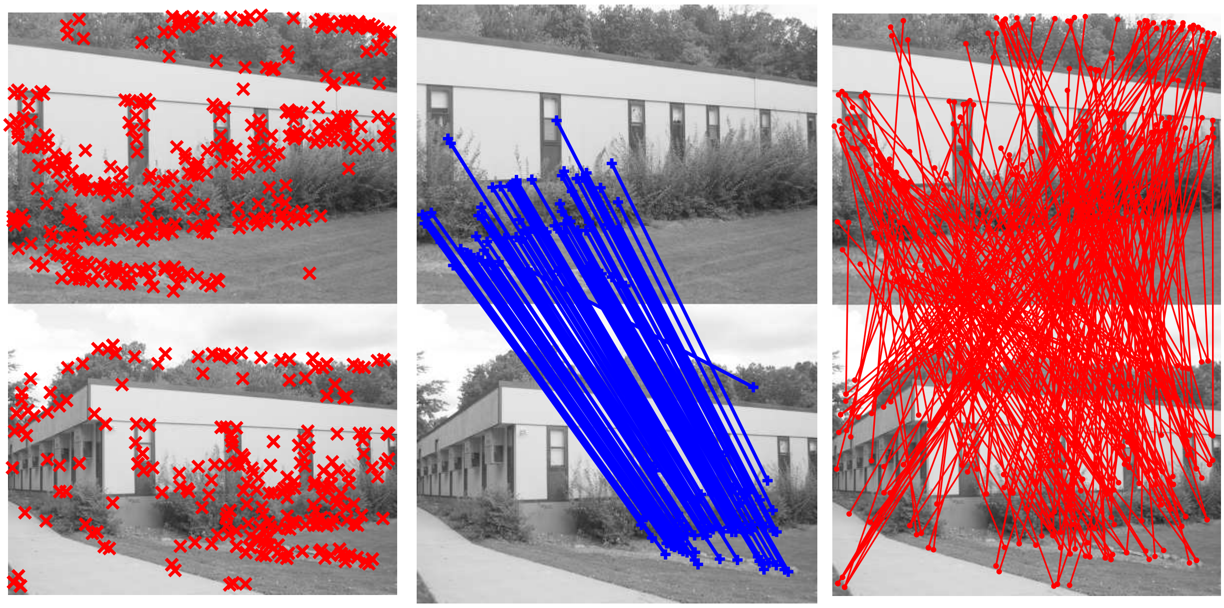
(a) correspondences. (b) identified inliers. (c) identified outliers.

Figure 6: 750 correspondences are initiated with around 50% outliers. 364 inliers are identified without false positive.



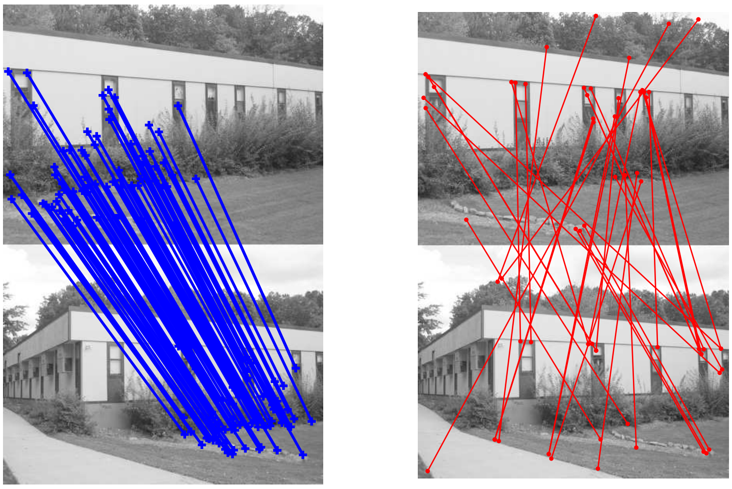
(a) correspondences. (b) identified inliers. (c) identified outliers.

Figure 7: 285 correspondences are initiated with about 70% outliers. 36 inliers are identified with only 3 false positives.



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

Figure 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.



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

Figure 9: Guided by distance between SIFT descriptors, 100 correspondences were selected, which contains about 30% outliers. 64 inliers were identified without false positive. See Figure 8 for original matches.