Nonparametric estimation of multiple structures with outliers

Wei Zhang and Jana Kosecka

Department of Computer Science, George Mason University,
4400 University Dr. Fairfax, VA 22030 USA
{wzhang2,kosecka}@cs.gmu.edu

Abstract. Common problem encountered in the analysis of dynamic scene is the problem of simultaneous estimation of the number of models and their parameters. This problem becomes difficult as the measurement noise in the data increases and the data are further corrupted by outliers. This is especially the case in a variety of motion estimation problems, where the displacement between the views is large and the process of establishing correspondences is difficult. In this paper we propose a novel nonparametric sampling based method for estimating the number of models and their parameters. The main novelty of the proposed method lies in the analysis of the distribution of residuals of individual data points with respect to the set of hypotheses, generated by a RANSAC-like sampling process. We will show that the modes of the residual distributions directly reveal the presence of multiple models and facilitate the recovery of the individual models, without making any assumptions about the distribution of the outliers or the noise process. The proposed approach is capable of handling data with a large fraction of outliers. Experiments with both synthetic data and image pairs related by different motion models are presented to demonstrate the effectiveness of the proposed approach.

1 Introduction and Related Work

In many computer vision estimation problems the measurements are frequently contaminated with outliers. Thus a robust estimation procedure is necessary to estimate the true model parameters. In practice, data can contain multiple structures (models), which makes the estimation even more difficult. In such case for each structure, data which belong to other structures are also outliers (pseudo outliers) in addition to the true outliers (gross outliers).

The problem of robust estimation received lot of attention in computer vision literature. Most works on robust estimation focus on the estimation of a single model and typically differ in their assumptions, efficiency and capability of handling different fractions of outliers. With the exception of a few, the problem of robust estimation of multiple models received notably smaller attention and several previously proposed methods were either natural extensions of the robust techniques used for single model estimation. They proposed to estimate individual models iteratively or focused more on the model selection issues.

In computer vision community the two most commonly used techniques for dealing with noisy data and outliers are Hough transform and RANdom SAmple Consensus (RANSAC) [1] algorithm. In Hough transform multiple models are revealed as multiple peaks in the parameter space. The localization of these peaks in multi-dimensional space becomes more difficult as the noise and the number of outliers grow. The RANSAC algorithm, initially introduced for robust estimation problems with a single model, has been extended to the multiple model scenario. The existing RANSAC approaches differ in the choice of the objective function used to evaluate each individual hypothesis. The two most commonly used criteria, which the objective function typically captures are: 1) the residuals of the inliers should be as small as possible and 2) the number of inliers should be as many as possible. In the standard RANSAC, the second criterion is applied and hypotheses are ranked by the number of data points within some error bound, *i.e.*, inliers. The hypothesis with the most inliers is then chosen as the model and the model parameters are re-estimated with its inliers . The need for predefined inlier threshold is disadvantageous. Recently in [2] traditional RANSAC has been augmented by automatic scale (threshold) selection used to disambiguate the inliers and outliers and the authors have shown that a significant percentage of outliers can be tolerated. In [3], the author pointed out that using RANSAC for simultaneous estimation of multiple motions requires dramatically more samples than that of single motion case. As a result, motions are usually estimated sequentially to save the computation. However, evaluation of the motions individually violates the assumption that the outliers to the first motion form a uniform distribution. In the presence of multiple models, the remaining models serve as pseudo outliers, which are clustered rather than uniformly distributed. In [4] authors pointed out that clustered outliers are more difficult to handle than scattered outliers. In the context of structure and motion estimation, in [5] the author proposed a strategy to deal with multiple models. The method for determining the number of models was an iterative one and all the models were considered independently. Recently a novel algebraic technique

was proposed in [6], which enables simultaneous recovery of a number of models, their dimensions and parameters, assuming that the models can be characterized as linear subspaces of possibly different dimensions. The applicability of the approach has not been explored in the presence of a larger number of outliers.

Outline. In this paper we present a novel robust nonparametric sampling based method for simultaneous estimation of number of models and model parameters. This goal is achieved by studying the distribution of residuals for each data point. The residuals are computed with respect to a number of hypotheses generated in the sampling stage. We demonstrate that the number of modes in the distribution reflects the number of models generating the data and show how to effectively estimate these modes. The presented approach is demonstrated and justified on synthetic data. Several experiments with estimating multiple motion models on real data are presented to validate the approach.

2 The proposed approach

The approach described here shares some features of the method proposed in [7], but differs in significant ways, which enable significant extensions to estimation of multiple models. In [7] the authors propose a novel MDPE estimator (Maximal Density Power Estimator), which selects a hypothesis, whose corresponding density of residuals is maximal, with the mean close to zero. This entails the use of nonparametric techniques for studying the distribution of residuals of all data points with respect to individual hypotheses. The number of models can not be determined in one complete run of RANSAC, since only the best hypothesis is selected by RANSAC. Schindler and Sutter [8] recently proposed a scheme that can estimate multiple models simultaneously. The work focuses more on the model selection issues and criteria, which best explain the data. The associated optimization problem which they formulate is an NP-hard combinatorial problem. Taboo-search is used to find an approximate solution.

Instead of considering the residuals of all the data points per hypothesis, we propose to analyze the distribution with respect to all the hypotheses for each data point. Subsequent analysis of this distribution enables us to estimate the number of models as well as the parameters of the correct hypothesis consistent with the data points. First, for the simplicity and clarity of the notation, we will demonstrate the technique on a simple line fitting problem. Later on we will present the applicability of the method to the problem of estimation of multiple motions and multiple 3D planar structures from correspondences between two views.

Let N be the number of data points $\mathbf{x}_i \in \Re^n$ corrupted by noise. The available measurements then are

$$\mathbf{x}_i = \tilde{\mathbf{x}}_i + \delta\mathbf{x} \quad i = 1, \dots N.$$

Suppose that these data points are generated by multiple linear (or possibly non-linear) models, with parameters $\mathbf{v}$, such that each $\mathbf{x}_i$ belongs to at least one model. In linear case this constraint can be expressed algebraically as

$$(\mathbf{v}_1^T \mathbf{x}_i) \dots (\mathbf{v}_j^T \mathbf{x}_i) = 0 \quad j = 1, \dots D$$

where D is the number of models. Our goal is to estimate the number of models D as well as their parameters in case the data points are noise and further corrupted by a significant portion of outliers.

In the manner similar to the RANSAC algorithm, in the first stage the initial set of hypotheses (values of parameters $\mathbf{v}_j$) is generated by selecting minimal subsets of data points needed to estimate the model parameters. Let M be the number of hypotheses obtained in the sampling stage $h_j; j = 1 \dots M$. Instead of studying the distribution of N residuals per hypothesis as in [7] when trying to determine the threshold for inlier classification, we propose to study the distribution of M residuals for each data point $\mathbf{x}_i$. We will show that this distribution reveals the presence of multiple models and further demonstrate how to estimate their number and their parameters.

The rationale behind this choice is the following: when many samples are drawn from data containing multiple models, for each model, there will be a subset of samples which consist of only points belonging to it (inliers). For instance suppose that we are given data generated by three models, where the percentage of inliers for each model is 33%. If one (minimal) sample needed to estimate a hypothesis comprised of 4 points, then the probability that the sample is outlier free for one model is $0.33^4 = 0.012$. Given 3000 samples, the expected number[1] of outlier free samples is $0.012 \times 3000 = 36$. Since the points used to calculate the hypotheses come from the same model, hypotheses parameters $\mathbf{v}_j$ estimated based on them will be close and will form a cluster in the hypothesis space. The clusters of hypotheses will have similar behavior with respect to a particular data point $\mathbf{x}_i$, in the sense that the residuals of $\mathbf{x}_i$ with respect to the cluster of h_j's will be similar. The samples which contain outliers would also generate hypotheses, whose residuals will be randomly distributed in the residual space. As a result, the distribution of residuals for each data point will have peaks (modes) corresponding to the clusters of hypotheses. For instance, Figure 1(c) shows that a residual distribution for a bi-modal data set has two strong peaks. The similar idea of search for clusters of hypotheses is also the basis of Randomized Hough Transform [9]. In that case however the search for clusters proceeds in often multidimensional parameter space as opposed to residual space and hence is known to suffer from typical shortcomings of Hough Transform methods (e.g. localization accuracy, resolution and efficiency).

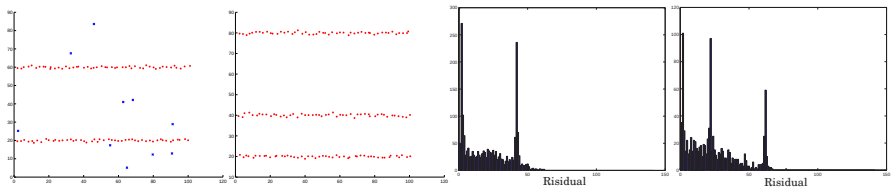
The observations outlined above give rise to the following four-step sampling based method for estimating of multiple models in the presence of a large number of outliers. In the following section we will demonstrate the individual steps of the proposed method in two simple examples. The first set of data points is

[1] The number of outlier free samples obeys a binomial distribution, the probability of success is the probability that a sample is outlier free.

Algorithm 1 Multiple Model Estimation

1. In the first stage M hypotheses are generated. The parameters of the hypotheses models are estimated from a minimal number of data points randomly drawn from the data.
2. For each data point $\mathbf{x}_i$, compute its residuals r_i^j for $j = 1 \dots M$ with respect to all the hypotheses.
3. The number of models D is estimated by determining the number of modes in residuals histograms of each data point. Final number is the median of all the estimates.
4. For each hypothesis, the correct cluster of model hypotheses is then identified.

generated by two parallel lines, each with 50 points corrupted by Gaussian noise $N(0, 0.5)$, 10 random points are added as outliers. The second set of data points contains three parallel lines, each with 50 points corrupted by Gaussian noise $N(0, 0.5)$. Figures 1(a) and 1(b) show the two configurations.



(a) (b) (c) (d)

Fig. 1. (a) and (b): the first and second data. (c) and (d): residual distribution of point from the first and second data.

2.1 Model hypothesis generation

Same as the standard RANSAC scheme, model hypotheses are computed using a minimal set of data points required to estimate the model[2]. The number of samples to be drawn is related to the percentage of outliers and the desired confidence of outlier free sample. The higher the outlier percentage, the more samples needed to ensure that a cluster of hypotheses will be generated. In RANSAC framework the number of required samples can be estimated theoretically assuming a known percentage of outliers ϵ and the desired probability ρ_s that the samples include at least one outlier free sample, because of the following relation:

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

[2] For instance, the minimal number is 2 for line fitting, and 4 for estimating inter-image homography.

where m is the number of samples and p is the number of points per sample (typically the minimal number of points needed to estimate the hypothesis). For the proposed approach, a set of outlier free samples is needed to form a mode (cluster) in the residual space. Therefore, we are interested in the probability ρ that at least K outlier free samples are included among m samples:

$$\rho = 1 - \sum_{i=0}^{K-1} p_i^m = 1 - \sum_{i=0}^{K-1} \binom{m}{i} (1-\epsilon)^{ip} (1-(1-\epsilon)^p)^{m-i} \tag{2}$$

where the term in the summation p_i^m is the probability that exactly i samples are outlier free in m samples. Equation 1 is a special case of Equation 2 for $K = 1$. In the standard RANSAC, Equation 1 is typically used to obtain a closed form solution for the required number of samples M:

$$M = \left\lceil \frac{ln(1-\rho)}{ln(1-(1-\epsilon)^p)} \right\rceil \tag{3}$$

needed for a desired confidence ρ. Using Equation 2 we can obtain the required number of samples, by computing how ρ changes while varying m for a desired K. Let's consider an example of estimating two homographies with the same number of supporting features with 20% gross outliers (i.e. 40% are valid for each motion), $p = 4$ in this case and $\epsilon = 0.6$ for each individual homography. Let assume that we need $K = 50$ outlier free samples, this is much more than enough to form a evident peak. We will see that the number of samples required would still be low for such a rigid requirement and with 2500 hypotheses samples, the probability would be:

$$\rho = 1 - \sum_{i=0}^{50} \binom{2500}{i} (1-0.6)^{4i} (1-(1-0.6)^4)^{2500-i} = 0.96$$

For varying m, the confidence ρ is shown in Table 1. Thus the required number of samples M can be obtained based on the table.

m	2000	2100	2200	2300	2400	2500	2600	2700
ρ	0.53	0.67	0.78	0.87	0.92	0.96	0.98	0.99

Table 1. The probability ρ for a given number of samples m.

For example given 2700 samples, the probability that both homographies have at least 50 outlier free samples would be $0.99 \times 0.99 = 0.9801$. In [3], Tordoff and Murray have shown that if RANSAC is used to estimate two motions simultaneously, the required number of samples increases dramatically over the single motion case. According to [3], to estimate two homographies in this example, the probability ρ_m that a desired sample is obtained in m samples is:

$$\rho_m = 1 - (1 - 0.5^4 0.5^4 0.8^8)^m$$

which can be simplified to be:

$$\rho_m = 1 - (1 - 0.4^4 0.4^4)^m.$$

The above expression captures the fact that a desired sample should contain 4 inliers of one homography and 4 inliers of the other homography simultaneously. In this case, 6000 samples are needed for 98% probability that a desired sample is included. On the other hand, the proposed algorithm can achieve the same probability with much less (2700) samples. The reduction of the number of samples is even more when the outlier percentage is higher.

One may argue that running RANSAC sequentially will be much more efficient. It has been observed that the theoretical number of samples is usually not sufficient for RANSAC. This is partly due to the fact that RANSAC needs samples that are not only outlier free but also well distributed. For example when doing line fitting($p = 2$), it's desirable that sampled points are well separated for better signal to noise ratio. Consequently, the actual number of samples for RANSAC can be 10 times as many as the theoretic number, observed in [3]. In our approach the theoretical number of samples is adequate, since we explicitly require a set of outlier free samples of a particular size. Also note here we are referring to the case when only one model is to be estimated by RANSAC. Sequential RANSAC procedures need to be run to estimate multiple models, let alone the fact that this is only an efficient approximation of the simultaneous estimation of multiple models.

Our experiments showed that doing RANSAC sequentially will not necessarily require less samples as demonstrated in Section 3.1. One reason is that for data with multiple structures, the set of samples will simultaneously contain outlier free samples for each structures. For example, assuming we have a data which consists of two equal lines and 50% outliers, if 300 samples ($p = 2$) are drawn from it, we can expect about 20 samples from the first line and 20 samples from the second line. The proposed method utilizes both sets of outlier free samples. While for RANSAC, only outlier free samples from one line are considered and the others are discarded. The second line has to be estimated by another run of RANSAC by generating a new set of samples. In fact, the number of samples were always on the order of 2000 throughout our experiments even though the data were quite challenging.

2.2 Residuals analysis

With M hypotheses generated, M residuals can be computed for each data point. For a general linear model the residual of a data point $\mathbf{x}_i$ with respect to the model $\mathbf{v}_j$ is $(r_i^j)^2 = (\mathbf{v}_i^{jT}\mathbf{x}_i)^2$. For line fitting examples the residuals are geometric distances between the points and the lines hypotheses. The residual of ith point with respect to the jth line is:

$$r_i^j = \frac{|a_j x_i + b_j y_i + c_j|}{\sqrt{a_j^2 + b_j^2}} \tag{4}$$

where $\mathbf{v}_j = [a_j, b_j, c_j]^T$ are the line parameters and $\mathbf{x}_i = [x_i, y_i]^T$ is the data point. Then the residual histogram of each data point denoted as f_i can be obtained for any point $\mathbf{x}_i, i = 1, \ldots, N$. As mentioned before, hypotheses estimated based on inliers to one model contribute to a peak (mode) in the histogram. This is demonstrated by the examples in Figure 1(c) and 1(d): there are two strong peaks in the residual histogram of one point in the first data set which contains two models. For a point in the second data set containing three models, three strong peaks stand out in its histogram of residuals.

One thing worth mentioning is that the number of residual distributions to be studied in our approach is N, whereas M residual distributions need to be studied in RANSAC framework [2]. When percentage of outliers is high (which is often the case in multi-modal data), $M \gg N$ to guarantee outlier free sample. Thus our approach is computationally more efficient in the residual histogram analysis stage. Furthermore the number of data points is usually limited, which might cause a poor estimate of the residual distribution per hypotheses as done in [2]. In our case the large number of hypotheses makes the approximation of residual distribution for each point feasible and more accurate.

2.3 Estimating the number of models

Since one mode corresponds to one model, the number of models can be estimated by identifying the number of modes in the residual histograms. While this is straightforward for the data in Figure 1(a) and 1(b), it's not easy for more noisy data containing many outliers. Figure 2(a) shows the residual histogram of a data point shown in Figure 4(a), where there are 3 models and 50% gross outliers. Identifying the modes which correspond to models requires careful treatment.

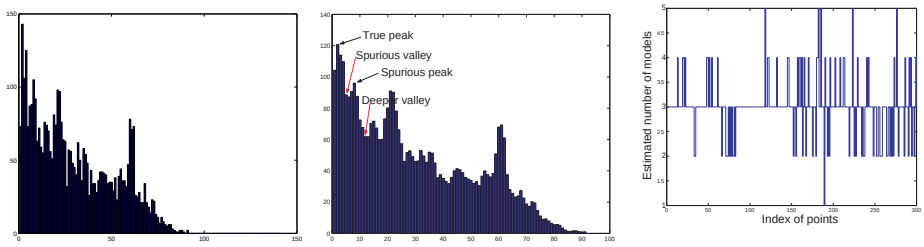
One possibility would be to employ one of the standard techniques for non-parametric probability density estimation, such as the Mean shift algorithm introduced to the vision community in [10]. The basic premise of the method is the estimation of the mean shift vector, which locally points in the direction of the maximum increase in the density and has been shown to converge to the modes. Both [2] and [11] pointed out some difficulties with the method in case of multi-modal data, as well as sensitivity of the mean shift algorithm with respect to the choice of bandwidth (size of the window) parameter. A tight bandwidth makes it very sensitive to local peaks, whereas correct modes would be missed with large bandwidth. This in particular is the case in our scenario, where the histograms contain many spurious peaks due to the presence of a large percentage of outliers. Since in our case we are limited to the analysis of 1D distributions of residuals, we have developed an alternative iterative procedure for detecting the models and disambiguating the correct modes from spurious ones. The mode detection method is summarized below:

The distinctness measure is defined as $\tau = f(\text{mode})/f(\text{shallow\_valley})$, where $f(i)$ is the histogram value of the ith bin. Let's look at the two left local modes of Figure 2(b), which is the smoothed result of Figure 2(a). Note that the true mode is not distinct enough from its valley, which is a spurious valley. Checking

Algorithm 2 Mode detection

1. In the first stage, the histogram is smoothed with a narrow window and local maxima (modes) and minima (valleys) are located.
2. Remove he spurious weak modes and valleys, so that only single local minimum valley is present between two modes and only one local maximum mode is presents between two valleys.
3. Choose the weakest unlabeled mode and measure its distinctness. If the mode is distinct, then it is labeled and added to the list of modes; otherwise it is marked as spurious and removed. If there are no more unlabelled modes, stop the procedure. Otherwise, go to step 2.

its distinctness directly would result in removing this correct mode. However, our approach guarantees that the spurious mode will be processed before the true peak. Since the spurious mode is not sufficiently distinct (τ less than some threshold T_τ) from its left (shallow) valley, it is removed in Step 3 of the procedure. Then the correct mode will obtain deeper valley after Step 2, enabling it to pass T_τ. Note that it is important that shallow valley is used for the comparison. The spurious modes close to the correct strong mode usually have deeper valleys of much smaller value. Only their shallow valleys reflect the fact that they are spurious modes.



(a) One residual histogram. (b) Peak identification. (c) Estimated # models.

Fig. 2. Identifying the number of models.

¿From each residual histogram f_i, we obtain an estimate d_i of the number of peaks and hence the number of models. Note that the residual histograms are different for different points and it's likely that d_i will be different for different i. Figure 2 plots the estimated $d_i, i = 1, \ldots, 300$ for each of the 300 data points in Figure 4(a). Most of the estimated numbers are equal to 3. The median of those numbers $d_m = \text{median}(d_i)$ provides a reliable estimate of the number of models.

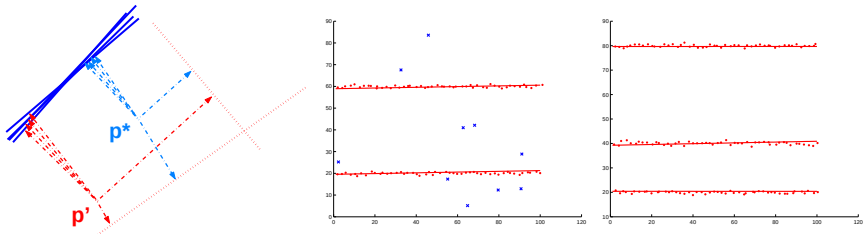
2.4 Finding the correct hypothesis and models parameters

Once the number of models has been obtained, we select a subset S of the data points. $S = \{\mathbf{x}_i | d_i = d_m\}$, which returned the correct number of models. Among them we select a point $\mathbf{x}_s$ whose histogram f_s has the strongest peaks

$$s = \arg\max_j \prod_{i=1}^{d_m} f_j(\text{peak}(i)) \tag{5}$$

where $f_j(\text{peak}(i))$ is the ith peak's magnitude of the residual histogram of jth point in S, $\mathbf{x}_s$ and f_s are then used to identify the correct models hypotheses.

For each identified mode, the corresponding hypothesis and consequently the model is determined as following: we first locate a subset of hypotheses whose residuals r_s^j correspond to the mode. We know that a cluster of hypotheses corresponding to a true model will be included in it, but it may happen that some additional random hypotheses also have the same residuals. Then, the problem is how to identify the good hypothesis from the chosen hypotheses subset. One possibility would be to apply a clustering method in the parameter space in the spirit of Hough Transform. This would result in a more efficient approach than Hough Transformation applied to the original problem, since only a subset of hypotheses needs to be checked. Yet we find a more efficient way, by searching for the clusters in the 1D residual space and by exploiting the distribution of residuals of another data point. Figure 3(a) illustrates the idea. The residuals (distances) of $\mathbf{x}_s$ and a set of line hypotheses are approximately the same, including correct hypotheses (solid lines colored blue) and spurious hypotheses (dotted lines colored red). To disambiguate them, we choose another random point $\mathbf{x}_i$, $i \neq s$ and study its residual distribution. Clearly, residuals of $\mathbf{x}_i$ will be different for the chosen hypotheses, but the clustered hypotheses will still have roughly the same residuals, thus forming a peak in the new residual distribution. The hypothesis which corresponds to the center of the peak will be selected as the model. The results of the synthetic examples are shown in Figure 3(b) and 3(c), respectively. Note that we don't need to identify inliers throughout the procedure, thus avoiding the need of inlier threshold.



(a) (b) (c)

Fig. 3. Identifying the model parameters.

3 Experiments

In order to assess the proposed method, we carried out various experiments. Line fitting was tested first, followed by motion estimation problem from two view correspondences, both with synthetic and real data.

3.1 Multiple Line Fitting

We carried out experiments on the line fitting problem with a number of data configurations, by varying the number of lines, percentage of outliers and noise level. Four experiments are shown in Figure 4. The image sizes are 100×100. The ith line has n_i data points, perturbed by Gaussian noise $N(0, \sigma^2)$. κ points are uniformly distributed within the image as outliers. Then we can compute the percentage of outliers for ith line (including gross and pseudo outliers), denoted as ϵ_i.

(a) Three parallel lines, $n_i = 50$, $\sigma = 1$, $\kappa = 150$; $\epsilon_i = 83.3\%$.
(b) Outlier form a cluster, $n_i = 50$, $\sigma = 5$, $\kappa = 50$; $\epsilon_i = 50\%$.
(c) 6-lines, $n_i = 25$, $\sigma = 1$, $\kappa = 50$; $\epsilon_i = 87.5\%$.
(d) 6-lines, $n_i = 25$, $\sigma = 0.3$, $\kappa = 50$; $\epsilon_i = 87.5\%$.

Our experiments showed that the method can tolerate a high level of outliers and a significant level of noise. For instance, $\epsilon_i = 87.5\%$ for one line in Figure 4(c). The noise standard deviation is large, 1% of image size for most tests. Only when data are rather complex (6 lines in the image), did our approach not succeed to fitting all the lines, yet 3 of them still got detected. When data points are less noisy, more lines can be detected. As Figure 4(d) shows, our approach correctly estimated the number of models and 5 lines were correctly estimated, when $\sigma = 0.3$. This is roughly equivalent to 2 pixel gaussian noise in a typical image of size 640. Another interesting observation is that our approach is fairly robust to a cluster of outliers, as Figure 4(b) shows. As people have already noticed [4], concentrated outliers are more difficult to handle than scattered outliers. According to the result of [7], existing robust estimators including RANSAC are likely to fail in this case. Figure 4(b) shows that the correct line can still be identified. Our approach predicted that there are two models in data, and detected one spurious line. This is actually not very surprising, since the cluster of outliers can be considered as a degenerate line.

We also want to mention our alternative experiments which used RANSAC sequentially to address the problem. Because the outlier ratios are high, each run of RANSAC required considerable number of samples. For example, in the case of Figure 4(c), even though 1000 samples were used for each run of RANSAC (3000 samples were used for estimating 3 lines), we still could not get consistent correct results. Figure 5(a) shows a typical result incorrect result, while the proposed method correctly returned 3 lines using 3000 samples. The inlier threshold was set optimally for the RANSAC procedure. As for the case of Figure 4(d), running RANSAC sequentially gave the result shown in Figure 5(b). Not all the

lines were correctly estimated. When outliers are clustered, sequential RANSAC returned wrong results even though the number of samples is the same as our approach. Figure 5(c) shows the result when the number of models was set to be 1. Figure 5(d) shows the result when the number of models was set to be 2. These instances were solved correctly by our approach.

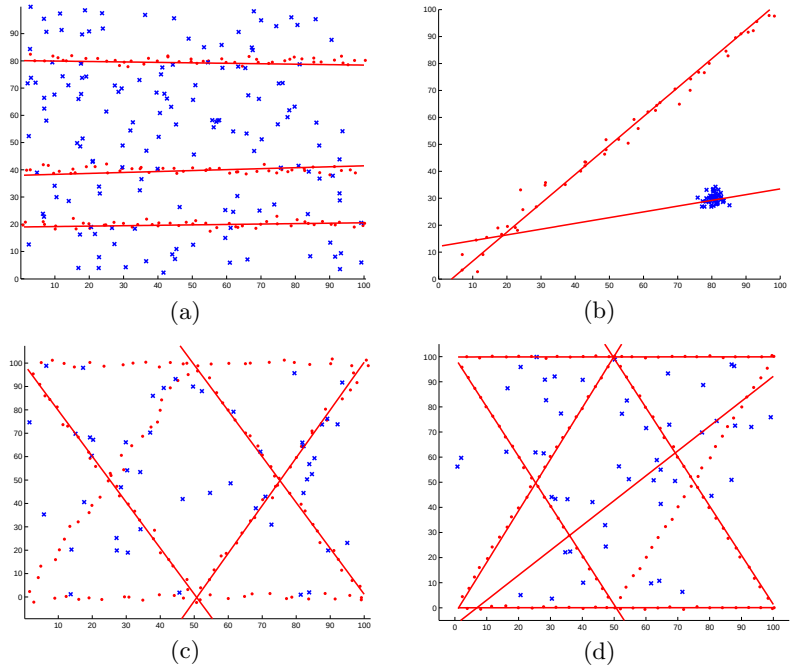


Fig. 4. The line fitting experiments, inliers are denoted as red '.', outliers are denoted as blue 'x'.

3.2 Two view correspondences

Synthetic data was tried first. The original data lie in 3D space, containing two planes, each with 40 points randomly distributed on that plane. Then they are projected into two views, with image sizes around 500. The points coordinates are corrupted by Gaussian noise of 0.5 pixels, and 20 outliers are randomly distributed in the image plane. As shown in Figure 6(b), both the number of homographies and their parameter are estimated correctly.

The approach was also applied to real images. In one experiment, we tried to identify planar surfaces in the image by estimating homographies. 60 correspondences belonging to two plane were manually selected. 40 random outliers were added. As Figure 7 shows, two planes are identified and their inliers are marked.

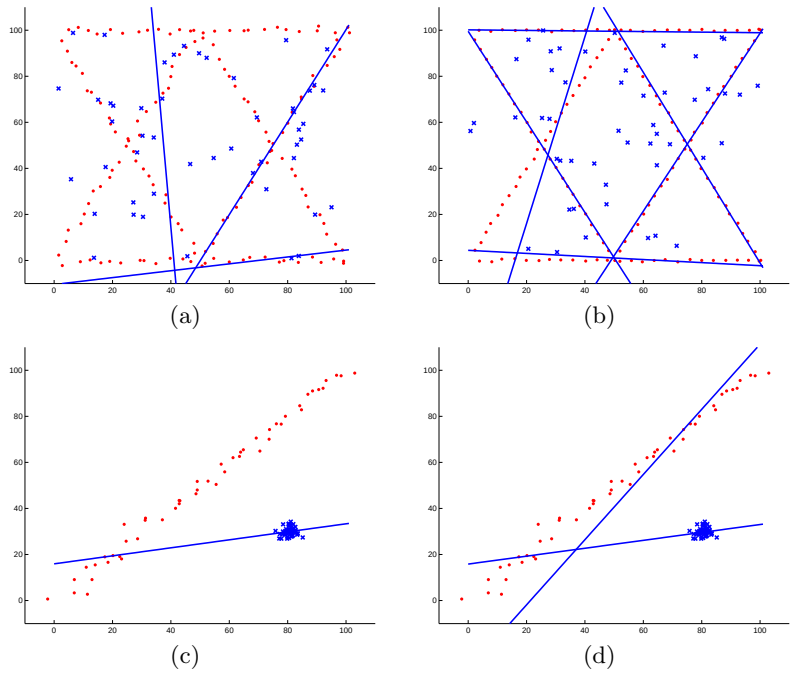
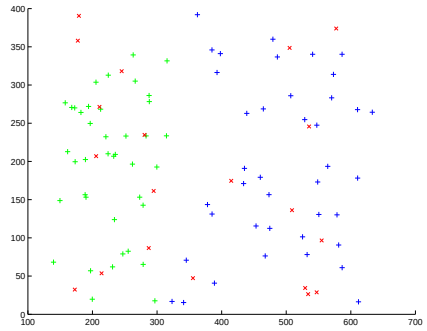
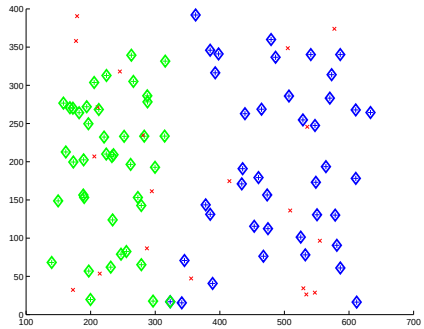


Fig. 5. The line fitting experiments using RANSAC sequentially.



(a) One view of original data. Data points from two planes are represented as '+', and colored blue and green respectively. The outlier points are represented as 'x' and colored red.

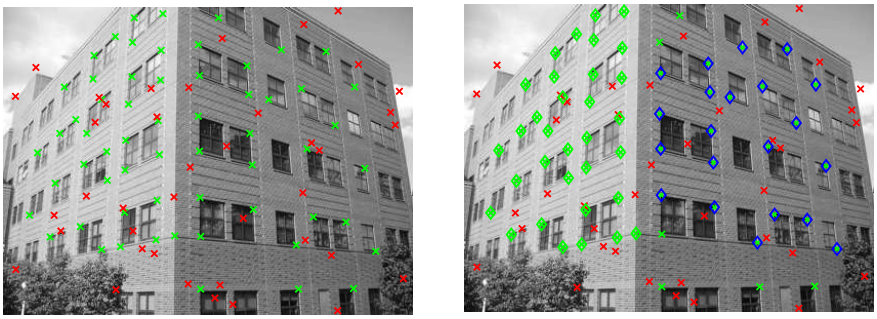


(b) Identified inliers to each model are denoted by '◊'. Only one data point close to the border of the two planes is labeled incorrectly.

Fig. 6. The experiment with homography model.

In another experiment, we tried motion segmentation for three sequences downloaded from http://www.suri.it.okayama-u.ac.jp/e-program-separate.html. Figure 8 shows the segmentation results using 2D translation or affine model. Both

the number of models and model parameters were correctly estimated for each sequence.



(a) One view of the image pair. Data points colored green. The outlier points are colored red.

(b) Identified inliers to each model are denoted by '◊', and colored blue and green, respectively.

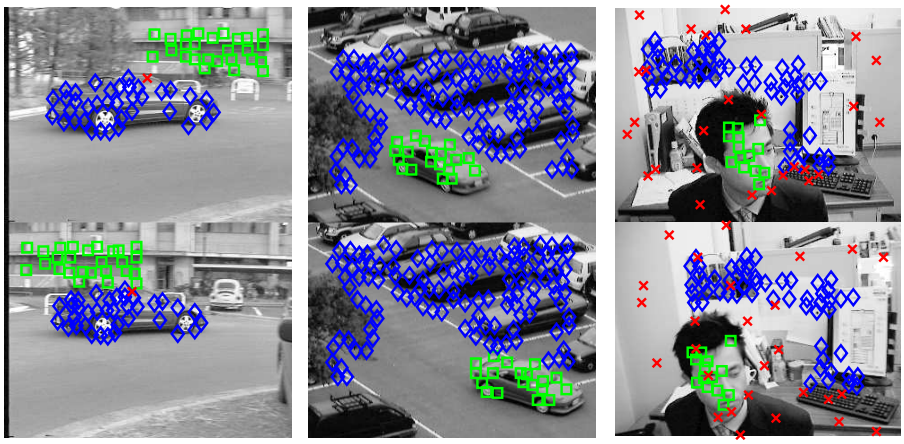
Fig. 7. The experiment with homography model. Two homographies are correctly estimated.

4 Conclusion

In this paper, we proposed a robust estimation scheme for multi-modal data with outliers. Base on the analysis of the residuals distribution per individual data points with respect to a set of hypotheses (generated by RANSAC-like sampling process), we can simultaneously estimate the number of models and parameters of each model. An iterative technique is developed to robustly identify the correct modes in the residual histogram, which is then used to determine the number of models. Model parameters are recovered from cluster in residual space instead of parameter space as done by Hough Transform, so the proposed approach will not suffer from the common difficulty of Hough Transform. Our approach was justified by extensive experiments on both synthetic and real data. Currently, we are investigating the structure and motion estimation problem with the proposed framework.

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: Comm. ACM,. (1981) 381–395
2. Wang, H., Suter, D.: Robust adaptive-scale parametric model estimation for computer vision. IEEE Trans. Pattern Anal. Mach. Intell. **26** (2004) 1459–1474



(a) (b) (c)

Fig. 8. Motion segmentation result. Identified inliers to each model are denoted by '◊' and '□', and colored blue and red, respectively. Identified outliers are denoted by red 'x'. (a) Affine segmentation of a car moving sequence. Note one of the correspondences is labeled as outlier because its position is not correct. (b) 2D translational segmentation a car leaving a parking lot. (c) 2D translational segmentation of head moving sequence. 20% random correspondences are added as outliers. The segmentation result is unaffected by the outliers.

3. Tordoff, B., Murray, D.W.: Guided sampling and consensus for motion estimation. In: ECCV (1). (2002) 82–98
4. Rocke, D., Woodruff, D.: Identification of outliers in multivariate data. Journal of the American Statistical Association **91** (1996) 1047–1061
5. Torr, P.: Geometric motion segmentation and model selection. Philosoplhical Transactions of the Royal Society of London **356** (1998) 1321–1340
6. Vidal, R., Ma, Y., Sastry, S.: Generalized principal component analysis (gpca). In: CVPR. (2003) 621–628
7. Wang, H., Suter, D.: MDPE: A very robust estimator for model fitting and range image segmentation. IJCV **59** (2004) 139–166
8. Schindler, K., Suter, D.: Two-view multibody structure and motion with outliers. In: CVPR'05. (2005)
9. Xu, L., Oja, E., Kultanen, P.: A new curve detection method: Randomized Hough Transform (RHT). Pattern Recogn. Lett. **11** (1990) 331–338
10. Comaniciu, D., Meer, P.: Mean shift analysis and applications. In: ICCV. (1999) 1197–1203
11. Chen, H., Meer, P., Tyler, D.: Robust regression for data with multiple structures. In: CVPR. (2001) 1069