

Robust Statistical Estimation and Segmentation of Multiple Subspaces

Allen Y. Yang, Shankar R. Rao, and Yi Ma *

Coordinated Science Laboratory, University of Illinois at Urbana-Champaign

1308 W. Main, Urbana, Illinois 61801

{yangyang,srrao,yima}@uiuc.edu

Abstract

We study the problem of estimating a mixed geometric model of multiple subspaces in the presence of a significant amount of outliers. The estimation of multiple subspaces is an important problem in computer vision, particularly for segmenting multiple motions in an image sequence. We first provide a comprehensive survey of robust statistical techniques in the literature, and identify three main approaches for detecting and rejecting outliers. Through a careful examination of these approaches, we propose and investigate three principled methods for robustly estimating mixed subspace models: random sample consensus, the influence function, and multivariate trimming. Using a benchmark synthetic experiment and a set of real image sequences, we conduct a thorough comparison of the three methods.

1. Introduction

The estimation of subspace models is an important problem in computer vision. For instance, in multiple-view geometry, we are often faced with the problem of recovering the parameters of a single motion in a dynamic scene from image features. Under most conventional camera models, such as orthographic, paraperspective, and affine models, one can show that image features associated with a rigid body motion lie on a subspace of dimension less than five [5, 16, 34]. Thus one can recover the motion parameters by estimating a linear subspace model. [28] further proposed a unified framework to represent the projective epipolar and homography relations using subspace models.

The problem becomes more complicated when there are multiple motions in the scene. Scenes with multiple motions are very common in many real-world applications. For example, a surveillance sequence of a parking lot may contain feature sets from multiple cars, each of which may undergo a different motion. In this situation, we are interested in the problem of segmenting image features to

different moving objects, which is equivalent to estimating multiple subspaces – one subspace for each motion. If the measurement of the image features is perturbed only by moderate *Gaussian* noise, we can use either *expectation-maximization* (EM) [19] or *K-Subspaces* [11] to iteratively estimate the subspaces. Alternatively, *generalized principal component analysis* (GPCA) [29] and its extensions [33] provide a noniterative solution, which also tolerates moderate data noise. Spectral clustering methods have also been used to heuristically segment samples in different subspaces [5, 16], but the results have been shown more sensitive to noise than the previous two approaches [27, 31, 34].

Unfortunately, though the previous methods can provide stable model estimates in the presence of moderate noise, they will fail when the data are contaminated by outliers. In imagery data, outliers often correspond to false features or features that are not from the objects of interest. A model estimation method is called “robust” if it tolerates a certain amount of outliers in the sample data.¹ In this paper, we provide a rigorous and thorough study of the subspace segmentation problem when outliers are present. More precisely, we are interested in the following problem:

Problem Formulation. Let $\mathcal{A} = V_1 \cup V_2 \cup \dots \cup V_n \subset \mathbb{R}^D$ be a union of n linear subspaces (i.e., all pass the origin) in general position² with possibly different dimensions d_1, d_2, \dots, d_n . Assume that a set of samples $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ are drawn from \mathcal{A} , and are contaminated by certain amount of noise and outliers. Given the number of the subspaces and their dimensions, estimate the underlying subspaces and segment the inlying data to their respective subspaces.

Notice that the problem becomes more difficult if the number of subspaces and their dimensions are unknown. In this case, one has to rely on certain model selection criteria [16, 21, 33] to balance the model complexity and the data fidelity. For brevity, we do not address the model selection issue in this paper.

Contributions. Various robust techniques have been developed to deal with the outlier issue in many computer vision

*This work was partially supported by UIUC ECE and CSL startup funds, NSF CAREER IIS-0347456, ONR YIP N00014-05-1-0633, and UIUC CSE Fellowships.

¹Some *stable* algorithms in the literature claim to be robust but in the sense of noise tolerance, e.g., [12].

²In this paper, *general position* means the dimensions of the intersections of all subspaces are the smallest possible.

applications. For work on estimating single subspace models, we refer to [10, 15, 23, 26]. However, to the best of our knowledge, only a few methods address the robustness issue for estimating multiple subspaces. The proposed solutions either highly rely on a good initialization [19], or assume that the subspaces have special properties that cannot be easily generalized to other situations, e.g., orthogonality, no intersections, or same dimensions [5, 16, 21, 25].

In this paper, we conduct a systematic study on the robust estimation of mixed subspace models. Under a very general assumption that a data set is drawn from a union of linear subspaces, we propose and investigate three principled approaches to robustly estimate the subspace parameters, namely, *random sample consensus* (RANSAC), *the influence function*, and *multivariate trimming*. Using a benchmark synthetic sample set and a set of real motion sequences, we conduct a thorough comparison of the three approaches. All MATLAB codes are available on our website: <http://perception.csl.uiuc.edu/gpca/>.

2. Robust Techniques: A Literature Review

Despite centuries of study, there is no universally accepted definition for “outliers.” Most approaches are based on one of the following assumptions:

1. Probability-based: Outliers are a set of *small-probability* samples with respect to the probability distribution in question. A data set is atypical if such samples constitute a significant portion of the data. Methods in this approach include *M-estimators* [3, 14] and its variation, *multivariate trimming* (MVT) [9].
2. Influence-based: Outliers are samples that have relatively *large influence* on the estimated model parameters [10, 15]. The influence of a sample is normally the difference between the model estimated with and without the sample.
3. Consensus-based: Outliers are samples that are *not consistent* with the model inferred from the remainder of the data. A measure of inconsistency is normally the error residue of the sample in question with respect to the model. Methods in this approach include Hough transform [1], RANSAC [7], and its many variations [4, 20–22, 25, 30].

In computer vision, various techniques have been derived based on these three assumptions. For instance, RANSAC was first used to estimate fundamental matrices [24], and was then extended to estimate multiple homography relations [2] and a mixture of epipolar and homography relations [21]. [17, 32] used M-estimators and MVT to robustify PCA. [15] also discussed how to robustify PCA using the influence function.

One important index of robust methods is the *breakdown point*, which is the minimal percentage of outliers in a data set that can cause arbitrarily large estimation error. It can be shown that the breakdown point for most probability-based

and influence-based methods is 50% [10, 23]. This drawback motivates the investigation of consensus-based methods. These techniques treat outliers as samples drawn from a model that is very different from the model of inliers. Therefore, although the outlier percentage may be greater than 50%, they may not result in a model with higher consensus than the inlier model.

The breakdown point also depends on the definition of the model. In the context of multiple subspaces, if one chooses to estimate and extract one subspace at a time, the inlying samples from all other subspaces together with the true outliers become the outliers for the subspace of interest. Therefore, one has to adopt a consensus-based method in searching for an estimator with a high breakdown point. On the other hand, if one chooses to treat the union of the subspaces as a single model, probability-based and influence-based approaches may also be applied to achieve good estimation, as long as the true outliers do not exceed 50% of the total sample points.

In Section 3, we compare two ways to estimate multiple subspaces using random sampling techniques. The first one estimates the union of the subspaces as a single model, and the other estimates one subspace at a time. Alternatively, GPCA and its variations estimate multiple subspaces by recovering a set of polynomials that vanish on the union of the subspaces. However, as we will explain later, all previous algorithms do not handle outliers well, i.e., the breakdown points are 0%. In Section 4, we propose a novel way to detect and eliminate outliers in estimating the vanishing polynomials based on traditional robust statistical techniques.

3. Subspace Segmentation via RANSAC

As we have mentioned, there are two ways to apply the random sampling scheme to estimate a mixed subspace model. Since in our problem, the number and the dimensions of the subspaces are given, one can either sample a larger subset to estimate the union of the subspaces (referred to as RANSAC-on-Union), or estimate and extract one subspace at a time (referred to as RANSAC-on-Subspaces). In the computer vision literature, the latter approach dominates most applications [2, 4, 21, 25, 30] because applying RANSAC on individual simple geometric models has been well studied, and the algorithm complexity is much lower.

To illustrate the dramatic difference of the complexity, suppose a set of 1800 valid samples is evenly drawn from three hyperplanes of dimension five in \mathbb{R}^6 , and the data set is further contaminated by 20% outliers. To estimate a single subspace model, we need five points (plus the origin), and with respect to a single subspace, the outlier percentage is 73.3%. To reach 95% confidence that one subset is outlier free, one needs to sample 2220 subsets.³ However, to estimate the three subspaces as a union, we need to sample 15 points, evenly partition the set into three subsets, and es-

³While the joint confidence of the whole model will be slightly less than 95%.

Table 1. RANSAC-on-Union applying to three models with 6% Gaussian noise and 24% outliers.

Model	(2, 2, 1) in \mathbb{R}^3	(4, 2, 2, 1) in \mathbb{R}^5	(5, 5, 5) in \mathbb{R}^6
Inlier Size	(200, 200, 100)	(400, 200, 200, 100)	(600, 600, 600)
Number of Samplings	5,000	1,000,000	10,000,000
Angle Error (degree)	1.5	3.5	10.2
Time Lapse	27s	5 hours	> 2 days

timate three subspace bases from the three subsets, respectively. To reach 95% confidence that one subset is outlier free, one needs to sample about 7.27 billion subsets.

However, our experiments show that RANSAC-on-Union can achieve good accuracy as long as enough iterations are provided. We implemented this method in MATLAB on a dual 2.7GHz G5 Apple workstation, and the results of three simulations are shown in Table 1. We use the average subspace angle error between the *a priori* model and the estimated one to measure the accuracy of the algorithm. The numbers of iterations shown in the table are the smallest ones in order to achieve reasonable estimation results, although they are much smaller than the theoretical ones with 95% confidence.

3.1. RANSAC on Individual Subspaces

To sequentially estimate multiple subspaces via RANSAC, most solutions assume that either the subspaces have the same dimensions or the data do not contain samples on the intersections of the subspaces. Special care has to be given when we deal with a more general situation. In fact, many complications arise when we try to apply RANSAC to a mixture of subspaces: 1. If one tries to find a higher-dimensional subspace first, the model may over-fit one or multiple lower-dimensional subspaces, and they are more likely to rank high in the consensus test. 2. If one tries to estimate a lower-dimensional subspace first, a subset from a higher-dimensional subspace or even the intersection of subspaces may likely win out first in the consensus test.

These types of ambiguities have been well known in computer vision as a potential problem of RANSAC. In multiple-view geometry, if both a planar object and a general 3-D object are present, a RANSAC process that searches for a fundamental matrix may overfit points from the planar object since a homography is a degenerate epipolar relation, which causes an erroneous estimation of the epipolar geometry [4]. However, in the general subspace segmentation problem, the situation is much more delicate, as multiple subspaces or samples at intersections may give high consensus to a subspace model in the presence of both noise and outliers. These “rare” situations are indeed very common in many applications. For instance, in *hybrid linear systems*, a single linear system may satisfy a subspace constraint of an arbitrary dimension, and the switching between multiple systems will generate output samples that are close to subspace intersections [13]. Figure 1 illustrates the complications in a simple toy example.

Recently, two modifications have been proposed to address this problem:

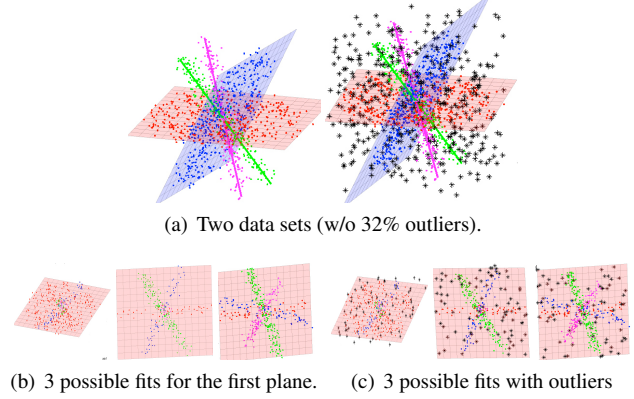


Figure 1. Possible segmentation results (in color) to fit the first plane model on samples drawn from 4 subspaces of dimensions (2, 2, 1, 1). The support of this 2-D model may come from samples on the true plane, or multiple degenerate line models. The degeneracy becomes more difficult to detect with outliers.

1. Starting with the highest-dimensional model, after a minimal sample set of the model achieves a high consensus, the algorithm further verifies if subsets of this minimal set result in high consensus of lower-dimensional models. If this is true, the algorithm will re-estimate the model from the remaining samples [4].
2. Alternatively, [8, 21] suggested a means to simultaneously apply RANSAC on multiple model hypotheses at a much larger scale, and use a model selection criterion to decide the number and types of models that best represent the mixed data.

In this section, we implement RANSAC-on-Subspaces using the first method.⁴ Its performance is demonstrated in Section 5. We observe that the algorithm highly relies on the degeneracy testing for subspaces of different dimensions. With more uniformly distributed outliers added in, the degeneracy testing becomes more inefficient (e.g., Figure 1), which leads to declined accuracy.

4. Robust GPCA

Given a set of sample points $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ drawn from a union of subspaces $\mathcal{A} = V_1 \cup V_2 \cup \dots \cup V_n$ in \mathbb{R}^D , GPCA seeks to simultaneously infer the subspaces and segment the data points to their closest subspaces. The key idea is to identify a set of polynomials

⁴For the second method, the model selection step via either minimum description length or maximum likelihood estimation inevitably introduces new heuristic parameters to the process. The algorithm complexity in over-sampling more models usually is also much higher.

$P = \{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})\}$ of degree n that vanish on (or fit) all the sample data. The way to estimate the polynomials is to use the *Veronese embedding* ν_n of degree n that embeds the data points \mathbf{x} into a higher-dimensional space [29]. The coefficients of each vanishing polynomial $f(\mathbf{x})$ is then in the left null space of the following embedded data matrix

$$L_n \doteq [\nu_n(\mathbf{x}_1), \nu_n(\mathbf{x}_2), \dots, \nu_n(\mathbf{x}_N)] \in \mathbb{R}^{M_n^{[D]} \times N}, \quad (1)$$

where $M_n^{[D]} = \binom{n+D-1}{D-1}$ is the number of monomials of degree n in D variables, and $m := \dim(\text{Null}(L_n))$ is uniquely determined by the Hilbert function of the subspaces [33]. Once the vanishing polynomials are found, the derivatives of the polynomials $DP = \{\frac{d}{d\mathbf{x}}f(\mathbf{x})\}$ at each sample point \mathbf{x}_i give the normal vectors to the subspace to which \mathbf{x}_i belongs, which give a segmentation of the whole sample set with respect to the subspaces. In the presence of noise, a voting scheme can be applied to improve the *stability* of the estimated subspaces [33].

However, it is known that the breakdown points of these algorithms are 0%, i.e., an add-in outlier with a large magnitude can arbitrarily perturb the estimated subspaces. The reason is that the coefficients of the vanishing polynomials $P = \{f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})\}$ correspond to the smallest eigenvectors of the data matrix L_n . A single outlier may arbitrarily change the eigenspace of L_n , which is estimated by PCA, and result in erroneous coefficients of the polynomials in P . Therefore, to eliminate the effect of outliers, we are essentially seeking a *robust PCA* method to estimate $\text{Null}(L_n)$ such that it is insensitive to outliers, or to reject outliers before estimating $\text{Null}(L_n)$.

In the robust statistics literature, there are two major approaches to robustify PCA with high-dimensional multivariate data, namely, the influence function and the robust covariance estimator [15]. In this section, we first discuss a simpler situation in which the percentage of outliers is known. We introduce two methods to robustly estimate the coefficients of a set of linearly independent vanishing polynomials by both the *influence function* and *multivariate trimming* (MVT). When the percentage of outliers is not given, we propose a method to estimate the percentage in Section 4.3.

4.1. The Influence Function

As we have noticed above, the null space $\text{Null}(L_n)$ is spanned by a set of eigenvectors $C = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m\}$, whose coefficients correspond to the coefficients of a set of linearly independent polynomials $f_i(\mathbf{x}) = \mathbf{c}_i^T \nu_n(\mathbf{x})$ of degree n that vanish on the valid sample data. The influence of a particular sample \mathbf{x}_i on $\{f_1, f_2, \dots, f_m\}$ is measured by the difference between the subspace $\text{Null}(L_n)$ estimated from all the N samples and the subspace $\text{Null}(L_n^{(i)})$ estimated from all but the i th sample:

$$I(\mathbf{x}_i) \doteq \langle \text{Null}(L_n), \text{Null}(L_n^{(i)}) \rangle, \quad (2)$$

where $\langle \cdot, \cdot \rangle$ denotes the subspace angle difference between two subspaces. All samples then can be sorted by their influence values, and the ones with the highest values will be rejected as “outliers”, and will not be used for the estimation of the eigensubspace (or the vanishing polynomials).

In robust statistics, $I(\mathbf{x}_i)$ is called the *sample influence function*. Notice that the complexity of the resulting algorithm is rather high. Suppose we have N samples, then we need to perform PCA $N + 1$ times in order to evaluate the influence values for the N samples. In light of this drawback, several formulae have been proposed to approximate the function, which are usually called *theoretical influence functions*. The reader may refer to [18] for more development about using theoretical influence functions to robustify GPCA.

4.2. Multivariate Trimming

If we treat the vectors $\{\nu_n(\mathbf{x}_i)\}$ in L_n as random samples, the problem of robustly estimating $C = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m\}$ becomes how to robustly estimate the covariance matrix of the random vector $\mathbf{u} = \nu_n(\mathbf{x})$. It is shown in [6] that, if both the valid samples and the outliers are of zero-mean Gaussian distribution and the covariance matrix of the outliers is a scaled version of that of the valid samples, the *Mahalanobis* distance:

$$d_i = \mathbf{u}_i^T \Sigma^{-1} \mathbf{u}_i \quad (3)$$

based on the empirical sample covariance $\Sigma = \frac{1}{N-1} \sum_{i=1}^N \mathbf{u}_i \mathbf{u}_i^T$ is a sufficient statistic for the optimal test that maximizes the probability of correct decision about the outlier (in the class of tests that are invariant under linear transformations). Thus, one can use d_i as a measure to down-weight or discard outlying samples while trying to estimate a robust sample covariance Σ .

Depending on the choice of the down-weighting schemes, many robust covariance estimators have been developed in the literature, among which is multivariate trimming (MVT). The convergence rate of MVT is the fastest among all methods of this kind, and its breakdown point is the percentage of samples trimmed from the data set.

The MVT method proceeds as follows. As the random vector $\nu_n(\mathbf{x})$ is not necessarily zero mean, we first obtain a robust estimate of the mean $\bar{\mathbf{u}}$ of the samples $\{\mathbf{u}_i = \nu_n(\mathbf{x}_i)\}$ (using techniques such as in [9]). We then need to specify a trimming parameter $\alpha\%$, which essentially is equivalent to the outlier percentage. To initialize the covariance matrix Σ_0 , all samples are sorted by the Euclidean distance $\|\mathbf{u}_i - \bar{\mathbf{u}}\|$, and Σ_0 is calculated as:

$$\Sigma_0 = \frac{1}{|U| - 1} \sum_{h \in U} (\mathbf{u}_h - \bar{\mathbf{u}})(\mathbf{u}_h - \bar{\mathbf{u}})^T, \quad (4)$$

where U is an index set of first $(100 - \alpha)\%$ samples with the smallest distance. In the k th iteration, the Mahalanobis distance of each sample, $d_k = (\mathbf{u}_i - \bar{\mathbf{u}})^T \Sigma_{k-1}^{-1} (\mathbf{u}_i - \bar{\mathbf{u}})$, is calculated, and Σ_k is again calculated using the set of

first $(100 - \alpha)\%$ samples with the smallest Mahalanobis distance. The iteration terminates when the difference between Σ_{k-1} and Σ_k is small enough.

To proceed with the rest of the GPCA algorithm, we treat the trimmed samples in the final iteration as outliers, and estimate $P = \{f_1(x), f_2(x), \dots, f_m(x)\}$ from the last m eigenvectors of the resulting covariance matrix.

Example 1 We apply both the influence function and MVT to robustify the GPCA-Voting algorithm [33] on the three synthetic models in Table 1 with various outlier percentages. The only heuristic parameter in GPCA-Voting, namely, the angle tolerance τ , is fixed at 0.3 rad, and the a priori outlier percentages are given to the algorithms. Figure 2 shows the results of the average angle errors. MVT performs much better than the influence function in the three data sets, particularly when the outlier percentage is high. The highest angle errors for MVT are all within 4 degree.

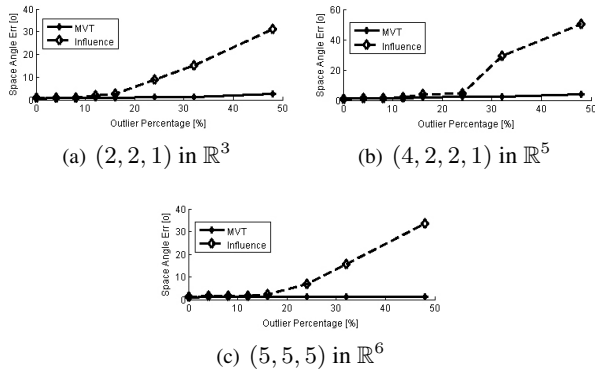


Figure 2. Average space angle errors of the robustified GPCA algorithm via the influence function and MVT.

Finally, we emphasize that although the distribution of the data is multi-modal of multiple subspaces in the original data space, the embedded data in the Veronese space become uni-modal as a single subspace model. One can show that any sample that perpendicular to $\text{Null}(L_n)$ indeed vanishes on the set of polynomials $P = \{f_1(x), f_2(x), \dots, f_m(x)\}$, and therefore is on the union of the subspaces in the original data space [18]. In MVT, the Mahalanobis distance is a sufficient statistic by assuming a Gaussian distribution of $\mathbf{u} = \nu_n(x)$, which is not true under the Veronese embedding. Nevertheless, this approximation of the distribution gives consistent good performance in our simulated and real experiments.

4.3. Estimating the Outlier Percentage

The above algorithms did not completely solve the outlier issue, since usually we do not know the outlier percentage for a given data set. In this subsection, we propose a means to estimate the outlier percentage. The percentage will be so determined that the GPCA algorithm returns a “good” mixed subspace model from the remaining sample

points. The main idea is to conduct the outlier rejection process multiple times under different rejection rates, and verify the “goodness” of the resulting models.

We first illustrate the basic ideas with an example. We randomly draw a set of sample points from three subspaces of dimensions $(2, 2, 1)$ in \mathbb{R}^3 with sample sizes $(200, 200, 100)$ and add 6% Gaussian noise. Then, the data are contaminated by 16% uniformly distributed outliers. We use MVT to trim out various percentages of samples ranging from 0% to 54%, and compute the maximal residual of the remaining samples with respect to the model estimated by GPCA. Figure 3 shows the plot of the maximal residual versus the rejection rate. The maximal sample residual reaches a plateau right after 7% rejection rate, and the residual decreases when the rejection rate increases. Figure 4 shows the segmentation results at rejection rate 7% and 38%, respectively.

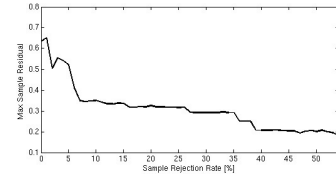


Figure 3. Maximal sample residual versus rejection rate.

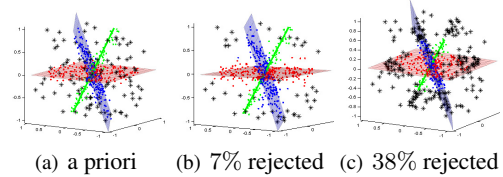


Figure 4. Subspace segmentation results.

In the experiment, although the 7% rejection rate is far less than the *a priori* 16% outlier percentage, the remaining outliers left in the sample set are nevertheless close to the subspaces (in terms of their residuals w.r.t. the estimated mixed model), and the resulting subspaces are close to the ground truth. We also see that MVT is moderately stable when the rejection rate is higher than the actual percentage of outliers. In this case, when the rejection rate is 38%, MVT trims out inlying samples that have relatively larger noise, which results in even a smaller maximal residual as shown in Figure 3. Therefore, one does not have to reject the exact *a priori* outlier percentage in order to obtain a good estimate of the model. In the presence of both noise and outliers, e.g., Figure 4(b), it is impossible and unnecessary to distinguish outliers that close to the subspaces from valid samples that have large noise.

Principle (Outlier Percentage Test). A good estimation of the outlier percentage can be determined by the influence of the outlier candidates with respect to the estimated (subspace) models. That is, further rejection from the data set only results in small changes in both the model parameters estimated and the fitting error.

This principle suggests two possible approaches for determining the rejection rate from the plot of the maximal sample residual:

1. The rejection rate can be determined by finding the first “knee point,” or equivalently the first “plateau,” in the residual plot (in the above example, at 7%).
2. The rejection rate can be determined by a pre-specified maximal residual threshold.

In practice, one may choose to use either approach based on the nature of the application. However, for the first approach, it is commonly agreed in the literature that a method that finds knee points and plateaus in a plot may not be robust if the data are noisy, since they are both related to the first-order derivatives of the plot. In addition, a well-shaped plateau may not exist in the residual plot at all if the *a priori* outlier percentage is small.

Therefore, in this paper, we determine the outlier percentage as the smallest one such that the maximal sample residual is smaller than a given residual threshold for several consecutive rejection rates, i.e., the residual “stabilizes.” The residual threshold can also be seen as the variance of the noise of the inlying data, which is similar to the same parameter in other robust statistical techniques, in particular RANSAC. Algorithm 1 gives an outline of the resulting algorithm. In practice, we find that three consecutive trials of 1% increments works well in both simulations and real experiments.⁵

Algorithm 1 (Robust GPCA).

Given a set of samples $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ in \mathbb{R}^D , a threshold τ for the subspace angle, and a residual threshold σ , fit n linear subspaces of codimensions c_1, c_2, \dots, c_n :

- 1: Set a maximal possible outlier percentage $M\%$.
 - 2: Normalize the data such that the max vector magnitude is 1.
 - 3: **for all** rejection rate $0 \leq r \leq M$ **do**
 - 4: $\mathbf{X}' \leftarrow$ removing $r\%$ samples from \mathbf{X} using MVT or Influence Function.
 - 5: Estimate the subspace bases $\{\hat{B}_1, \hat{B}_2, \dots, \hat{B}_n\}$ by applying GPCA to \mathbf{X}' with parameters τ and c_1, c_2, \dots, c_n .
 - 6: Maximal residual $\sigma_{\max} \leftarrow \max_{\mathbf{x} \in \mathbf{X}'} \min_k \|\mathbf{x} - \hat{B}_k \hat{B}_k^T \mathbf{x}\|$.
 - 7: **if** σ_{\max} is consistently smaller than σ **then**
 - 8: $B_k \leftarrow \hat{B}_k$ for $k = 1, 2, \dots, n$. Break.
 - 9: **end if**
 - 10: **end for**
 - 11: **if** $\sigma_{\max} > \sigma$ **then**
 - 12: ERROR: the given σ is too small.
 - 13: **else**
 - 14: Label $\mathbf{x} \in \mathbf{X}$ as an inlier if $\min_k \|\mathbf{x} - B_k B_k^T \mathbf{x}\| < \sigma$.
 - 15: Segment the inlying samples to their respective subspaces.
 - 16: **end if**
-

⁵It implicitly requires that the samples from any single subspace shall be more than 3% of the total data.

5. Experiments

5.1. Simulations

We test RANSAC-on-Subspaces and RGPCA on the three subspace models in Table 1 with various outlier percentages. Each data set has a maximal magnitude of one. For RANSAC, the boundary threshold is fixed at 0.1. For RGPCA using either MVT or Influence, the residual threshold σ is fixed at 0.04, and the angle threshold τ is fixed at 0.3 rad. Figure 5 shows the results of the average angle error. Table 2 shows the average time of the three algorithms with 24% outliers.

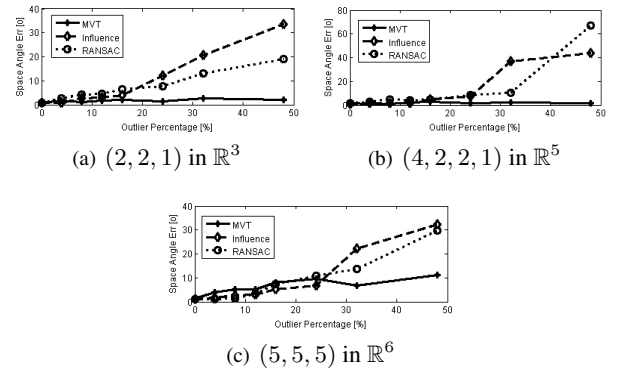


Figure 5. Average space angle errors of RANSAC and RGPCA (50 trials at each percentage).

Table 2. Average time of RANSAC and RGPCA with 24% outliers.

Arrangement	(2,2,1) in \mathbb{R}^3	(4,2,2,1) in \mathbb{R}^5	(5,5,5) in \mathbb{R}^6
RANSAC	44s	5.1m	3.4m
MVT	46s	23m	8m
Influence	3m	58m	146m

We compare the performance of the two RGPCA algorithms with that of RANSAC. Both MVT and Influence methods run slower than RANSAC, particularly when the number and dimensions of the subspaces are high. With respect to the accuracy, MVT gives the best overall estimation of the mixed subspace models on all three synthetic data sets. The subspace angle errors for MVT in (2, 2, 1) case and (4, 2, 2, 1) case are both within 2 degree with up to 50% outliers. For (5, 5, 5) case, the worst angle error for MVT is 11 degree. Finally, RANSAC outperforms the influence function method in most cases.

5.2. Motion Segmentation

We also apply RANSAC-on-Subspaces and RGPCA on a set of real motion sequences for segmenting features that belong to different moving objects. To fairly evaluate the performance of the methods as generic subspace segmentation algorithms, no information other than the coordinates of the tracked features is used in the experiment.

Suppose N 3-D points p_1, p_2, \dots, p_N on a rigid object are tracked from F consecutive frames of a motion sequence. Denote $\mathbf{m}_{ij} \in \mathbb{R}^2$ as the image coordinates of the i th point in the j th frame. Then it can be shown in [16] that all vectors of the form

$$\mathbf{x}_i = [\mathbf{m}_{i1}^T, \mathbf{m}_{i2}^T, \dots, \mathbf{m}_{iF}^T]^T \in \mathbb{R}^{2F}, \quad i = 1, 2, \dots, N,$$

under affine camera projection lie in a 4-dimensional subspace. Furthermore, if the object is a planar structure, or a 3-D structure undergoing a planar motion, the \mathbf{x}_i 's shall lie in a 3-dimensional subspace. We first project the stacked vectors $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ onto a 5-dimensional space using PCA. Then let the algorithms estimate multiple 3-D or 4-D subspaces with the number of subspaces and their dimensions given.

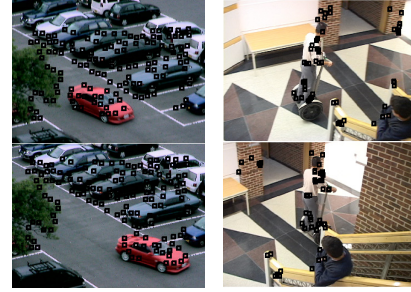
Four motion sequences shown in Figure 6 are used for testing. We separate the sequences into two categories. For the first two sequences, the objects all undergo rigid-body motions, and the cameras are far away from the scene. Therefore, the affine camera model is valid. For the last two sequences, the affine model is not strictly satisfied: In Sequence C, the camera is very close to the man, and the head motion is not a rigid-body motion; in Sequence D, the camera is also close to the scene, and partial occlusion of the objects causes false tracked features. For such sequences, the affine camera model becomes an approximation, and hence the stacked vectors $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ satisfy more complex models than 3-D or 4-D subspaces. Nevertheless, we command RANSAC and RGPCA to robustly fit subspaces to the data. We also use Sequence C and D to test the segmentation result on subspaces of different dimensions. For each sequence, the planar background is modeled as a 3-dimensional subspace, and the foreground objects are modeled as 4-dimensional subspaces.

Figure 7 shows the segmentation result. All parameters are tuned to achieve the best segmentation results. We can see that the algorithms perform reasonably well on the four sequences, considering no other imagery information is used to optimize the segmentation. The two RGPCA algorithms perform slightly better on Sequence B and D, and RANSAC performs better on Sequence C.

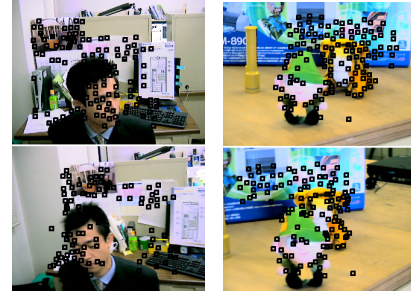
6. Conclusions

In this paper, we have examined three major robust statistical techniques to estimate multiple subspaces in the presence of moderate data noise and outliers. For random sampling techniques, our experiments have shown that RANSAC-on-Union requires a very large number of iterations to achieve good estimation results, which is not practical for most applications. RANSAC-on-Subspaces that sequentially extracts subspaces can handle subspaces of different dimensions by adding a degeneracy testing step. However, the testing may become inefficient when the outliers are abundant in the data set.

For robust GPCA, our experiments have shown that MVT provides better overall performance than the influence



(a) Sequence A: 140 points in 27 frames. (b) Sequence B: 53 points in 26 frames.



(c) Sequence C: 107 points in 81 frames. (d) Sequence D: 126 points in 81 frames.

Figure 6. The first and last frames of the four sequences.

function when the number and dimensions of the subspaces are moderate. However, two common drawbacks of both RGPCA algorithms are: 1. They do not provide a convenient estimate of the outlier percentage, because the outliers must be rejected in the Veronese embedded space. 2. The Veronese embedding causes RGPCA to not scale as well as RANSAC-on-Subspaces in terms of computational complexity when the subspace dimensions are high (e.g., > 10).

Taking account of other alternative solutions mentioned in the paper, e.g., using RANSAC to over-estimate a larger set of single subspace models (possibly via MCMC), it seems that no method can outperform others in all situations. Based on the nature of the data and the application, the reader is recommended to wisely choose or extend an appropriate method discussed in this paper.

Acknowledgment: We want to thank the anonymous reviewers and the program chairs for their valuable suggestions to improve the implementation of RANSAC and RGPCA.

References

- [1] D. Ballard. Generalizing the Hough transform to detect arbitrary patterns. *Pattern Recognition*, 13(2):111–122, 1981.
- [2] A. Bartoli. Piecewise planar segmentation for automatic scene modeling. In *CVPR*, 2001.
- [3] N. Campbell. Robust procedures in multivariate analysis I: robust covariance estimation. *Applied Statistics*, 29(3):231–237, 1980.

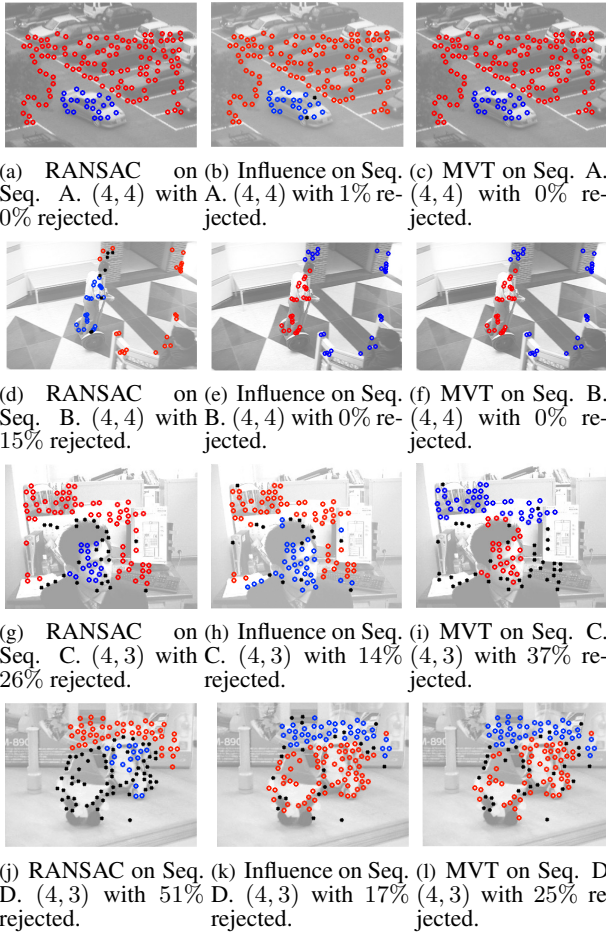


Figure 7. The segmentation result (in color). The black asterisks denote the outliers.

- [4] O. Chum, T. Werner, and J. Matas. Two-view geometry estimation unaffected by a dominant plane. In *CVPR*, 2005.
- [5] J. Costeira and T. Kanade. A multibody factorization method of independently moving objects. *IJCV*, 29(3):159–179, 1998.
- [6] T. Ferguson. On the rejection of outliers. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1961.
- [7] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24:381–85, 1981.
- [8] D. Forsyth, J. Haddon, and S. Ioffe. The joy of sampling. *IJCV*, 41:109–134, 2001.
- [9] R. Gnanadesikan and J. Kettenring. Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, 28(1):81–124, 1972.
- [10] F. Hampel, E. Ronchetti, P. Rousseeuw, and W. Stahel. *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 1986.
- [11] J. Ho, M. Yang, J. Lim, K. Lee, and D. Kriegman. Clustering appearances of objects under varying illumination conditions. In *CVPR*, 2003.

- [12] K. Huang, Y. Ma, and R. Vidal. Minimum Effective Dimension for Mixtures of Subspaces: A Robust GPCA Algorithm and Its Applications. In *CVPR*, pages 631–638, 2004.
- [13] K. Huang, A. Wagner, and Y. Ma. Identification of hybrid linear time-invariant systems via subspace embedding and segmentation. In *IEEE Conference on Decision and Control*, 2004.
- [14] P. Huber. *Robust statistics*. John Wiley & Sons, 1981.
- [15] I. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 2nd edition, 2002.
- [16] K. Kanatani. Motion segmentation by subspace separation: Model selection and reliability evaluation. *International Journal of Image and Graphics*, 2(2):179–197, 2003.
- [17] F. D. la Torre and M. J. Black. Robust principal component analysis for computer vision. In *ICCV*, pages 362–369, 2001.
- [18] Y. Ma, A. Yang, H. Derksen, and R. Fossum. Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *Submitted to SIAM Review*, 2006.
- [19] G. McLachlan and D. Peel. *Finite mixture models*. John Wiley & Sons, 2001.
- [20] P. Rousseeuw. Least median of squares regression. *Journal of American Statistics Association*, 79:871–880, 1984.
- [21] K. Schindler and D. Suter. Two-view multibody structure-and-motion with outliers. In *CVPR*, 2005.
- [22] C. Stewart. MINPRAN: a new robust estimator for computer vision. *PAMI*, 17(10):925–938, 1995.
- [23] C. Stewart. Robust parameter estimation in computer vision. *SIAM Review*, 41(3):513–537, 1999.
- [24] P. Torr. Geometric motion segmentation and model selection. *Phil. Trans. Royal Society of London*, 356(1740):1321–1340, 1998.
- [25] P. Torr and C. Davidson. IMPSAC: synthesis of importance sampling and random sample consensus. *PAMI*, 25(3):354–364, 2003.
- [26] P. Torr and D. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *IJCV*, 24(3):271–300, 1997.
- [27] R. Vidal and R. Hartley. Motion segmentation with missing data using PowerFactorization and GPCA. In *CVPR*, 2004.
- [28] R. Vidal and Y. Ma. A unified algebraic approach to 2-D and 3-D motion segmentation. In *ECCV*, 2004.
- [29] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (GPCA). *PAMI*, 27(12):1–15, 2005.
- [30] H. Wang and D. Suter. MDPE: a very robust estimator for model fitting and range image segmentation. *IJCV*, 59(2):139–166, 2004.
- [31] Y. Wu, Z. Zhang, T. Huang, and J. Lin. Multibody grouping via orthogonal subspace decomposition. In *CVPR*, pages 252–257, 2001.
- [32] L. Xu and A. Yuille. Robust principal component analysis by self-organizing rules based on statistical physics approach. *PAMI*, 6(1):131–143, 1995.
- [33] A. Y. Yang, S. Rao, A. Wagner, R. Fossum, and Y. Ma. Hilbert functions and applications to the estimation of subspace arrangements. In *ICCV*, 2005.
- [34] L. Zelnik-Manor and M. Irani. Degeneracies, dependencies and their implications in multi-body and multi-sequence factorization. In *CVPR*, 2003.