MINPRAN: A New Robust Estimator for Computer Vision

Charles V. Stewart, *Member, IEEE*

***Abstract*—MINPRAN is a new robust estimator capable of finding good fits in data sets containing more than 50% outliers. Unlike other techniques that handle large outlier percentages, MINPRAN does not rely on a known error bound for the good data. Instead, it assumes the bad data are randomly distributed within the dynamic range of the sensor. Based on this, MINPRAN uses random sampling to search for the fit and the inliers to the fit that are least likely to have occurred randomly. It runs in time $O(N^2 + SN \log N)$, where S is the number of random samples and N is the number of data points. We demonstrate analytically that MINPRAN distinguished good fits to random data and MINPRAN finds accurate fits and nearly the correct number of inliers, regardless of the percentage of true inliers. We confirm MINPRAN's properties experimentally on synthetic data and show it compares favorably to least median of squares. Finally, we apply MINPRAN to fitting planar surface patches and eliminating outliers in range data taken from complicated scenes.**

***Index Terms*—Surface reconstruction, robust estimation, range data, parameter estimation, outliers.**

I. Introduction

Robust vision techniques are needed to accurately fit parametrized functions to intensity, edge, or range data while ignoring gross errors ("outliers") in the data [15], [23], [33]. Since outliers are typical of both real-world sensors and low-level vision algorithms, robust techniques become more important as computer vision systems move from controlled laboratory settings to real applications. Already, a variety of robust techniques have been used in computer vision, some developed within the vision field [6], [8], [10], [11], [16], [34], others borrowed from statistics [3], [12], [15], [21], [25], [27].

Robust techniques may be characterized by their "breakdown point," the highest fraction of arbitrarily bad data that can be tolerated without these data being able to completely corrupt a fit [21]. For example, least sum of squares has a breakdown point of 0 because one arbitrarily bad point can completely corrupt a fit, regardless of the number of good points. On the other hand, least median of squares [20] has a breakdown point of 0.5 because up to half the data may be bad without altering the fit. However, for a variety of reasons, illustrated in Fig. 1, robust vision techniques often require a higher breakdown point:

- When more than half the data are bad, the correct fit may still be readily apparent.
- When the data arise from multiple surfaces, the points belonging to one surface are essentially outliers ("pseudo-outliers") with respect to the other surface(s), and fewer than 50% of the points may belong to any one surface.
- More than one surface may overlap the same image coordinates. This can occur if one surface is partially transparent, such as a wire fence, or when data from multiple views are merged. Here, the underlying surfaces can not be modeled as a single piece-wise continuous function of the image coordinates, and fewer than 50% of the data may arise from one surface.

A good robust estimator should find accurate fits in each of these cases. Furthermore, a good robust estimator should not "hallucinate" when there is no correct fit, as in Fig. 1d.

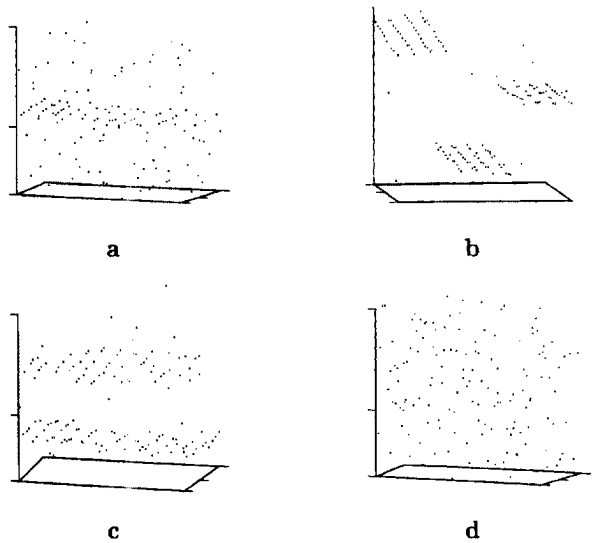


Fig. 1. Four example plots of synthetic depth data from 12 × 12 image regions. For each data set, the sensor viewpoint is from below the plot. Plot (a) contains 60% outliers. Plot (b) contains three different surfaces in a small region. Plot (c) contains two different surfaces that completely overlap in the region. In each case, none of the surfaces contain more than 50% of the points in the region, yet all correct fits are readily apparent. Plot (d) contains all random points and no correct fit is visible.

Rousseeuw argues the theoretical maximum breakdown point is 0.5 because if more than half of the data are bad they may "conspire" to look better than the correct fit [21]. If, however, we assume such a conspiracy is unlikely, we should be able to tolerate higher percentages of bad data, at least in a probabilistic sense. This implies we need to make some assumptions about the distribution of bad data. Hough transform techniques [2], [10] and Ransac techniques [8] can accept fits involving fewer than half of the data points by making an explicit assumption about the good data and an implicit assump-

Manuscript received Sept. 17, 1993; revised Nov. 16, 1994.

The author is with the Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180-3590; e-mail: stewart@cs.rpi.edu.

IEEECS Log Number P95119.

tion about the bad data. First, these techniques assume the good data are within a known distance (inlier bound) of the correct fit. Additionally, in order to distinguish good fits from random fits, they must implicitly assume the bad data are uniformly distributed [8], [9], [19]. In contrast, our new robust estimator surpasses the 0.5 breakdown point solely by assuming the bad data are uniformly distributed.[1] From this assumption we derive a criterion function to measure the probability a configuration of points near a hypothesized fit could have occurred randomly. Using this criterion function, we search a space of hypothesized fits, identifying the least random fit and its collection of inliers. We call the resulting estimator "MINPRAN" (MINimize the Probability of RANdomness).

Among existing robust techniques, MINPRAN is most similar to least median of squares (LMS) [15], [20], [25], which has a 0.5 breakdown point and (like MINPRAN) does not require a known inlier bound. Hence, we demonstrate MINPRAN's strengths by contrasting it with LMS.

- In finding the best fit, MINPRAN finds nearly the correct number of inliers, even when fewer than 50% of the points are inliers to any one fit. In fact, fits with an arbitrarily low inlier percentage (> 0) may be found, as long as the bad data are random and the good data are close enough to the correct fit.
- Since it identifies and uses essentially all of the inliers, MINPRAN produces more accurate fits than LMS, including more accurate estimates of the variance in the data.
- MINPRAN does not "hallucinate" fits when there is no correct fit, and with a slight modification to its basic search strategy, MINPRAN can find multiple fits when warranted by the data. LMS always finds exactly one fit for each set of data.

In this paper, we derive and analyze the MINPRAN estimator, presenting a theoretical analysis (backed by simulation results) of MINPRAN's ability to find good fits, to avoid hallucinating, and to surpass the 0.5 breakdown point. We then generalize MINPRAN's randomness model to arbitrary outlier distributions. Finally, by modifying the basic estimator to reliably find multiple fits in a data set, we develop a robust technique to fit surface patches to range data taken from complicated scenes. Note that during the theoretical discussion in most of the paper, we describe the data input to the estimator as a set of points. Only in Section VII, when we apply MINPRAN to range data, do we refer to images and the image regions to which the estimator is actually applied.

II. Probability of Randomness Criterion Function[2]

As the basis of our approach, we assume the dynamic range of the "sensor" (the source of the data) is known, and the bad data values are uniformly distributed within this dynamic range. Knowing the dynamic range of the sensor can be as simple as knowing the range of gray levels in an intensity image, the disparity range in stereo data, or the depth interval to which a range sensor is tuned. (We extend the model to non-uniform outlier distributions in Section VI.) In assuming the outliers are uniform within the dynamic range, we assume there is no known bias in the outlier data. From this assumption, we derive a criterion function based on the probability a given fit and its set of inliers could be due to purely random data.

Suppose a set of N points is drawn from a uniform distribution of z values in the range Z_{min} to Z_{max}, and consider a curve (or surface) ϕ and a distance r. Assuming $\phi \pm r$ is entirely within the range Z_{min} to Z_{max} (as in Fig. 2), the probability at least k points randomly fall within the range $\phi \pm r$ is given by a simple binomial distribution,

$$\sum_{i=k}^{N} \binom{N}{i} \left(\frac{r}{Z_0}\right)^i \left(1-\frac{r}{Z_o}\right)^{N-i}, \tag{1}$$

where $Z_0 = (Z_{max} - Z_{min})/2$. Defining the function $\mathcal{F}(r, k, N)$ to be this probability, we will use $\mathcal{F}$ as the basis for a criterion function that evaluates hypothesized fits to a data set. Before showing how we do this, we state two important properties of $\mathcal{F}$ which follow directly from the binomial distribution and its equivalence to the incomplete beta function [1]:

$$\begin{aligned}\mathcal{F}(r, k, N) &= \sum_{i=k}^{N} \binom{N}{i} \left(\frac{r}{Z_o}\right)^i \left(1-\frac{r}{Z_0}\right)^{N-i} \\ &= \frac{N!}{(k-1)!\,(N-k)!} \int_0^{r/Z_0} t^{k-1}(1-t)^{N-k}\,dt.\end{aligned} \tag{2}$$

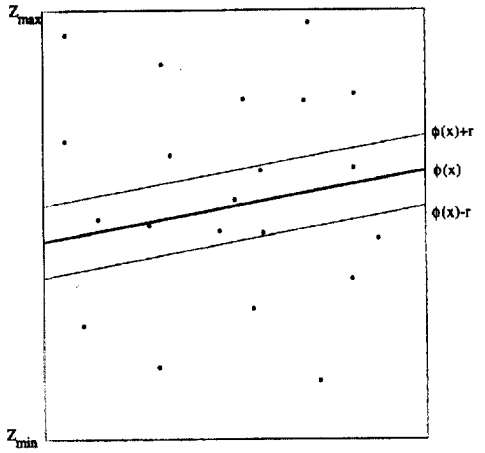


Fig. 2. If the data are uniformly distributed in the range Z_{min} to Z_{max}, then the probability that at least k points are within $\pm r$ of $\phi(x)$ is given by a binomial sum.

LEMMA 1. *$\mathcal{F}(r, k, N) > \mathcal{F}(r, j, N)$, for k, j, $1 \le k < j \le N$ and for fixed r, $0 < r < Z_0$.*

This states that the probability $\mathcal{F}$ decreases monotonically as the number of points in $\phi \pm r$ increases. It follows directly from (1) since increasing the number of point S in $\phi \pm r$ from k to j removes terms from the binomial summation.

1. Later in the paper, we will relax this assumption to allow any outlier distribution.

2. The notation has changed since earlier published versions of the work [29], [30], [31], [32].

LEMMA 2. *$\mathcal{F}(r, k, N) < \mathcal{F}(s, k, N)$, for $0 \le r < s \le Z_0$ and for fixed k, $1 \le k \le N$.*

This lemma states that for a fixed number points in $\phi \pm r$, the probability $\mathcal{F}$ increases monotonically as the distance r increases. It follows directly from (2) because increasing from r to s only increases the upper limit of integration, and since the integrand is strictly positive, this increases the value of the integral.

To derive our criterion function, let there be N data points and consider ϕ as a hypothesized fit to the data. Let r define an "inlier bound" to ϕ: points falling within $\phi \pm r$ are labeled inliers; points falling outside $\phi \pm r$ are labeled outliers. Furthermore, let $k_{\phi,r}$ be the number of data points within $\phi \pm r$, i.e. the number of points labeled inliers. (We use subscripts on k since it is completely determined by ϕ and r.) Then, $\mathcal{F}(r, k_{\phi,r}, N)$ gives the probability there could be least $k_{\phi,r}$ inliers within the inlier bound r of ϕ if all N points were random. If this probability is extremely low, it is unlikely all the points are really from a uniform distribution and much more likely at least some of them correspond to a curve (or surface, in three dimensions). Therefore, we use $\mathcal{F}$ to rate the "randomness" of a hypothesized fit to a data set, choosing as its representative value the minimum of $\mathcal{F}$ over all possible inlier bounds r:

$$\mathcal{H}(\phi, N) = \min_r \mathcal{F}(r, k_{\phi,r}, N)$$

$\mathcal{H}(\phi, N)$ is our "probability of randomness" criterion function. We search over the range of r to minimize $\mathcal{F}$ since we assume neither a fixed inlier bound (unlike Hough transforms and Ransac) nor a fixed number of inliers (unlike LMS).

Although r is continuous, calculating $\mathcal{H}(\phi, N)$ does not require examining all r, since $\mathcal{F}(r, k_{\phi,r}, N)$ has at most N local minima. This is established formally in the following lemma.

LEMMA 3. *For any hypothesized fit, ϕ, to N data points, there are at most N local minima of $\mathcal{F}(r, k_{\phi,r}, N)$. These local minima occur at the absolute residuals of the N points relative to ϕ.*

PROOF. Let r be an inlier bound for ϕ and let $k_{\phi,r}$ be the number of points within $\phi \pm r$. (Since $\mathcal{F}(r, 0, N) = 1$ for all $r < Z_0$, $\mathcal{F}(r, 0, N)$ cannot be a local minimum, so we do not consider values of r where $k_{\phi,r} = 0$.) Then, if we examine the (absolute) residuals of the N data points, there must be exactly $k_{\phi,r}$ residuals less than or equal to r. Hence, letting $k = k_{\phi,r}$ and writing the absolute residuals in non-decreasing order, $r_1, \ldots, r_N$, we must have $r_k \le r < r_{k+1}$. If $r > r_k$ then according to Lemma 2, $\mathcal{F}(r_k, k, N) < \mathcal{F}(s, k, N) < \mathcal{F}(r, k, N)$, for $r_k < s < r$, and there is no local minimum at r. Therefore, the local minimum of $\mathcal{F}(r, k_{\phi,r}, N)$ can only occur when $r = r_k$ for some k. □

This lemma implies if $r_{\phi,i}$, $1 \le i \le N$, are the fit residuals of the N points relative to ϕ, then

$$\mathcal{H}(\phi, N) = \min_i \mathcal{F}(r_{\phi,i}, i, N). \tag{3}$$

We take this as the final definition of the probability of randomness criterion function. (In cases where it is clear the residuals are relative to a particular fit, we drop the subscript ϕ and write the residuals as r_i.) The goal of the MINPRAN algorithm will be to minimize $\mathcal{H}(\phi, N)$ over a space of possible fits to a given set of N data points. Associated with the minimum will be both an inlier bound and a set of inliers. When the minimum of $\mathcal{H}$ is small enough, we will take the minimizing fit and its set of inliers as representing a curve or surface in the data (otherwise we say there is no curve or surface in the data).

Before describing MINPRAN, we make two concluding observations about $\mathcal{H}$. First, since the binomial summation defining $\mathcal{F}$ may be computed in time $O(N)$ (using a recurrence relation to generate the $O(N)$ terms of the summation), $\mathcal{H}(\phi, N)$ may be computed in $O(N^2)$ time by calculating and sorting the residuals, $r_{\phi,i}$, and then evaluating $\mathcal{F}(r_{\phi,i}, i, N)$ for each i. Second, we would obtain exactly the same criterion function if we assumed the fit residuals of the outliers were uniformly distributed in the range $[0 \ldots Z_0]$. Since this is only slightly different from what we actually assumed (it is equivalent when ϕ is in the center of the depth range), we will use this assumption in several stages of our analysis of MINPRAN.

III. THE MINPRAN ALGORITHM

Given a set of data points, we need an algorithm to find the fit to the data minimizing our probability of randomness criterion function, $\mathcal{H}$. The fit may be a linear function, such as a line for two-dimensional data or a plane for three-dimensional data, or it may be any other parameterized function. (In general, fits are of the form $f(\vec{x}, \vec{a}) = 0$, where $\vec{x}$ is the vector of model variables and $\vec{a}$ is the vector of model parameters.) By extending an $O(N^2 \log N)$ algorithm for least median of squares (LMS) regression [26], we have obtained an $O(N^3)$ time algorithm to find the line minimizing $\mathcal{H}$ in two dimensions. We hypothesize the planar fit minimizing $\mathcal{H}$ in three dimensions can be found in $O(N^4)$. Because these times are impractical for computer vision applications, we employ a random sampling technique [8], [15], [19], [20] to find an approximate minimum.

Assuming p points are required to completely instantiate a fit and there are $N' = N + p$ data points[3], MINPRAN chooses S distinct, but not necessarily disjoint, random subsets of p points from the data, and finds the fit to each random subset to form S hypothesized fits, $\phi_1, \ldots, \phi_S$. (See Roth and Levine[19] for a discussion of function fitting procedures in this context.) MINPRAN selects the fit minimizing $\mathcal{H}(\phi_j, N)$ as the "best fit".

As discussed above, evaluating $\mathcal{H}(\phi_j, N)$ for each hypothesized fit, ϕ_j, requires $O(N^2)$ time, yielding an $O(SN^2)$ algorithm to find best fit among the S fits tested. To improve upon this, observe the computation is dominated by the SN evaluations of $\mathcal{F}(1)$, once for each (ordered) residual to each fit. Fortunately, as stated in the following theorem, only N of these evaluations are necessary.

3. We use the notation $N' = N + p$ points because the p points chosen to instantiate a fit will automatically have 0 residuals relative to the fit; thus, MINPRAN considers only the residuals of the remaining N points in evaluating the criterion function H. This maintains consistency with other discussions throughout the paper.

THEOREM 1. *For a set of N data points and S fits, $\phi_1, \ldots \phi_S$, to these data points,*

$$\min_j \mathcal{H}(\phi_j, N) = \min_i \mathcal{F}(r_i^*, i, N),$$

where $r_i^* = \min_j\{r_{\phi_j,i}\}$ and $r_{\phi_j,i}$ is the ith ordered residual relative to ϕ_j.

PROOF. Let $r_{\phi_j,1}, \ldots, r_{\phi_j,N}$ be the residuals of the N points relative to ϕ_j. Then, according to the definition of $\mathcal{H}$ in (3),

$$\min_j \mathcal{H}(\phi_j, N) = \min_j\left\{\min_i \mathcal{F}(r_{\phi_j,i}, i, N)\right\}$$
$$= \min_i\left\{\min_j \mathcal{F}(r_{\phi_j,i}, i, N)\right\}$$

Let $r_i^* = \min_j\{r_{\phi_j,i}\}$, i.e., r_i^* is the smallest ith residual across the S fits. Then, because $\mathcal{F}(r, i, N)$ increases monotonically with r for fixed i (Lemma 2), $\min_j \mathcal{F}(r_{\phi_j,i}, i, N) = \mathcal{F}(r_i^*, i, N)$. Substituting this into the expression for $\min_j \mathcal{H}(\phi_j, N)$ gives the result. □

Thus, to find the best fit, MINPRAN computes the N residuals to each fit, compares the residuals between fits to find r_i^*, $i = 1, \ldots, N$, and evaluates $\mathcal{F}(r_i^*, i, N)$ for each i. Let ϕ^*, i^* and r^* be the fit, the number of inlier residuals, and the inlier bound producing the minimum. Computationally, calculating the N residuals to each of the S fits and sorting each set of residuals requires $O(SN \log N)$ time, finding the r_i^* values requires $O(SN)$ time, and the N evaluations of $\mathcal{F}$ require $O(N^2)$ time. This gives $O(SN \log N + N^2)$ time to find ϕ^*.

If $\mathcal{F}(r^*, i^*, N) < \mathcal{F}_0$, where $\mathcal{F}_0$ is a "randomness threshold" derived below, then ϕ^* is accepted as a correct fit. Otherwise, no fit is accepted. A final least-squares fit involving the $p + i^*$ inliers (the p points used to instantiate ϕ^* plus the i^* inlier residuals) to ϕ^* produces a more accurate estimate of the model parameters than the initial estimate using just p points, and estimates the noise variance, σ^2, of the data.

MINPRAN's random sampling process may be repeated to find more than one acceptable fit to the data (see Fig. 1, for example). In doing so, the $p + i^*$ inliers to ϕ^* are excluded from the random sampling and residual calculation processes. (This assumes correct fits include disjoint subsets of the data as inliers.) $\mathcal{H}(\phi, N)$ is still calculated assuming there are N residuals, however, to subject all fits to the same criteria. Fitting may continue as long as enough points remain and the minimum value of $\mathcal{H}$ resulting from the random sampling process is less than $\mathcal{F}_0$, which is computed once for each data set.

The two remaining issues in defining MINPRAN are the calculation of S, the number of random samples required, and the calculation of $\mathcal{F}_0$, the threshold used to distinguish good fits from fits to purely random data.

A. Number of Samples

S, the number of random samples required, may be determined from a user-specified minimum probability that at least one sample is chosen containing p points from a correct fit. Here, we generalize techniques for finding S when a single correct fit is in the data [15], [19], [20], to allow more than one correct fit (more than one function or surface from which the data arose). Doing this requires several parameters to be specified by the user. These parameters, which do not otherwise influence the algorithm, are the estimated maximum fraction of true outliers, x_0 (a "true" outlier is a point that does not belong to any correct fit), the minimum number of points to be allowed in a fit, m_0, and the estimated maximum number of correct fits, n_f. Based on these parameters, with N data points, let $b = \lfloor x_0 N \rfloor$ be the maximum number of true outliers and let $M = N - b$ be the minimum number of points that are inliers to various fits.

To derive S, suppose the data points arise from n_f surfaces, with m_i points, $1 \le i \le n_f$, on each surface, where $m_i \ge p$ and $\sum m_i \ge M$. Assume for the moment M mod $n_f = 0$ and let $m = M/n_f$. (In general, $m = \lfloor M/n_f \rfloor$.) The probability a random sample of p points is "good"—consists of all points from the same surface—is simply $\hat{q} = \sum_{i=1}^{n_f} C(m_i, p) / C(N, p)$, where $C(m_i, p)$ and $C(N, p)$ are binomial coefficients. It is straightforward to show q's minimum, which is the worst case, occurs when $m_i = M/n_f = m$, for all i, so we set $q = n_f C(m, p)/C(N, p)$. The probability at least one sample is "good" in S samples is, $P_g = 1 - (1 - q)^S$. If a minimum value of P_g is specified by the user, then the minimum number of samples required is

$$S = \left\lceil \frac{\log(1 - P_g)}{\log(1 - q)} \right\rceil. \tag{4}$$

In practice, there are several important refinements to this:

1) If $M/n_f < m_0$ then n_f and m_0 are inconsistent. Assuming m_0 is more significant, reassign $n_f = \lfloor M/m_0 \rfloor$ and $m = \lfloor M/n_f \rfloor$.
2) In recomputing S to find an additional fit, assuming no outliers were part of the fit accepted as correct, b stays constant, n_f is decremented, M is reduced by the number of accepted inliers, and m and q are then recomputed. The role of m_0 is important here, because if m_0 is small, then after the initial fit, a small initial outlier percentage can cause S to become extremely large. For example, if $m_0 = 3$, $p = 3$, $P_g = 0.99$, $n_f = 2$, $x_0 = 0.2$, and $N = 100$, then $b = 20$ and, initially, $S = 36$. If the initial fit has 75 inliers, then the new value of m is 5, and S will be 1,057. By setting $m_0 = 15$, after the initial fit $m = 15$ and $S = 21$.
3) If n_f is reduced to 0, then set $m = \max(m_0, N - b)$, and $q = C(m, p)/C(N, p)$.
4) Finally, set a lower bound (= 15) on the minimum number of samples allowed.

Examples values of S computed from (4) are shown in Table I. Several observations are apparent. First, as seen in the first three rows of the table, low percentages of true outliers require relatively few samples, even when multiple correct fits are expected. Second, as the percentage of true outliers increases, the number of samples required increases quickly (e.g., compare rows 4 and 5). Thus, although we will show

MINPRAN can find correct fits for large outlier percentages, the cost of finding them is relatively high. Finally, the effect of the number of points in a sample, p, is dramatic. For example, in switching from linear ($p = 2$, row 6) to planar ($p = 3$, row 2) to quadratic fits ($p = 6$, row 7) the number of samples increases dramatically.

TABLE I
NUMBER OF SAMPLES REQUIRED FOR VARIOUS PARAMETER VALUES. FOR EACH OF THE ENTRIES, $N = 50$ AND $m_0 = 10$.

Max Surfaces n_f	Sample Size p	Outlier Fraction x_0	Samples Required $P_g = .95$	Samples Required $P_g = .99$
3	3	0.1	42	64
2	3	0.1	18	27
2	3	0.3	42	65
1	3	0.3	8	12
1	3	0.6	60	91
2	2	0.1	7	10
2	6	0.1	318	489

The number of samples does not vary significantly with N unless N is quite small.

B. Randomness Threshold $\mathcal{F}_0$

MINPRAN accepts the best fit from S samples as correct if $\mathcal{F}(r^*, i^*, N) < \mathcal{F}_0$. $\mathcal{F}_0$ is a threshold based on the probability P_0 that the best fit to N uniformly distributed outliers is less than $\mathcal{F}_0$. (Intuitively, P_0 is the probability MINPRAN will hallucinate a fit where there is none.) Thus, for a user defined value of P_0 (e.g., $P_0 = 0.05$) we establish our threshold value $\mathcal{F}_0$.

For a given P_0, our goal is to find the value $\mathcal{F}_0$ such that from a set of points entirely composed of uniformly distributed outliers

$$Prob\left(\min_{1 \le j \le S} \mathcal{H}(\phi_j, N) < \mathcal{F}_0 \right) = P_0$$

or, equivalently,

$$Prob\left(\min_{1 \le j \le S, 1 \le i \le N} \mathcal{F}(r_{\phi_j, i}, i, N) < \mathcal{F}_0 \right) = P_0 \qquad (5)$$

To make the analysis feasible, we assume the S fits and their residuals are independent. Strictly speaking, this assumption is not correct, since the point set is the same for all fits. It is reasonable for relatively small values of S, however, and as we will see in the experimental results, it makes MINPRAN more conservative in accepting fits.

Equation (5) is solved using the following steps:

1) For a single fit ϕ, with ordered absolute residuals $r_{\phi,1}, \ldots, r_{\phi,N}$, and for any given value of $\mathcal{F}_0$, find the probability that $\min_i \mathcal{F}(r_{\phi,i}, i, N) > \mathcal{F}_0$. Call this probability $f(N, \mathcal{F}_0)$.
2) Use $f(N, \mathcal{F}_0)$ to find the probability that over S random samples, $\min_{j,i} \mathcal{F}(r_{\phi_j,i}, i, N) < \mathcal{F}_0$.
3) For a user specified value of this probability, P_0, search over the values of $\mathcal{F}_0$ to find the value satisfying (5).

B.1. Calculating $f(N, \mathcal{F}_0)$

Suppose the N residuals to a given fit are each drawn at random from a uniform distribution in the range [0 ... 1]. (Without loss of generality, for the remainder of this analysis we assume $Z_0 = 1$. To reflect this we refer to the residuals as p_i, and in special cases, f_i, rather than r_i. We also drop ϕ from the subscripting.) Let $p_1, \ldots, p_N$ be the residuals in non-decreasing order. For a given $\mathcal{F}_0$, our goal is to calculate

$$f(N, \mathcal{F}_0) = Prob\left(\min_i \mathcal{F}(p_i, i, N) > \mathcal{F}_0 \right).$$

In order for $\min_i \mathcal{F}(p_i, i, N) > \mathcal{F}_0$, we must have $\mathcal{F}(p_i, i, N) > F_0$ for each i. Since $\mathcal{F}$ increases monotonically with p_i for fixed i (see Lemma 2), there is a unique value, f_i, which can be found by bisection search, such that $\mathcal{F}(f_i, i, N) = \mathcal{F}_0$. Then, for our given set of residuals, $p_1, \ldots, p_N$, $\mathcal{F}(p_i, i, N) > \mathcal{F}_0$ if and only if $p_i > f_i$. Therefore,

$$f(N, \mathcal{F}_0) = Prob(\forall i, p_i > f_i).$$

For any given set of N residuals, in order for $p_i > f_i$ for all i, there must be no residuals in the range $0 \ldots f_1$, at most 1 residual in the range $0 \ldots f_2$, at most 2 residuals in the range $0 \ldots f_3$, etc. See Fig. 3. To state these constraints more succinctly, define $f_0 = 0$ and $f_{N+1} = 1$, and let c_i, $0 \le i \le N$, denote the number of residuals in the range $(f_i \ldots f_{i+1}]$. Based on these c_i counts, the constraints become,

$$\sum_{j=0}^{i} c_j \le i, \text{ for } 0 \le i < N, \text{ and } \sum_{j=0}^{N} c_j = N. \qquad (6)$$

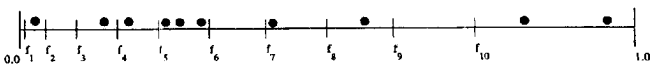


Fig. 3. An example set of residuals and f_i values for $N = 10$. The dots represent the residual values p_i, and each f_i value is such that $\mathcal{F}(f_i, i, N) = \mathcal{F}_0 = 0.05$. In this case there are six residuals between 0 and f_6, so $p_6 < f_6$ and $\mathcal{F}(p_6, 6, 10) < \mathcal{F}_0$. On the other hand, for $j \neq 6$, there are fewer than j residuals between 0 and f_j, so $p_j > f_j$ and $\mathcal{F}(p_j, j, 10) > F_0$.

To determine the probability of a random set of residuals satisfying these constraints, observe that because the residuals are uniformly distributed, the probability any particular residual is in the range $(f_i \ldots f_{i+1}]$ is simply Δf_i, where $\Delta f_i = f_{i+1} - f_i$. (For the remainder of this section we will consider the residuals prior to ordering.) Then, because the unordered residuals are independent, the probability c_i particular residuals are in the range f_i to f_{i+1} is $(\Delta f_i)^{c_i}$. Based on this, the probability of a given configuration of residual counts, $c_0, \ldots, c_N$ is

$$\frac{N!}{c_0!c_1!\ldots c_N!} \prod_{i=0}^{N} (\Delta f_i)^{c_i}$$

and, if C is the set of all configurations satisfying (6), we have

$$f(N, \mathcal{F}_0) = \sum_{(c_0, \ldots, c_N)} \frac{N!}{c_0!c_1!\ldots c_N!} \prod_{i=0}^{N} (\Delta f_i)^{c_i}. \qquad (7)$$

Thus, based on properties of $\mathcal{F}$, we have transformed the probability that $\mathcal{F}$ is greater than a threshold into an interval counting problem. The resulting equation can be evaluated in $O(N^3)$ time using a dynamic programming algorithm. (See [31] for details.)

B.2. Finding $\mathcal{F}_0$

Making the simplifying assumption that $\mathcal{F}$ is independent for each sample, the probability each of the S samples has $\min_i \mathcal{F}(p_i, i, N) > \mathcal{F}_0$ is just $(f(N, \mathcal{F}_0))^S$. Therefore, the desired probability at least one sample has a minimum less than $\mathcal{F}_0$ is

$$P_0 = 1 - (f(N, \mathcal{F}_0))^S. \quad (8)$$

Equation (5) can now be solved. Rearranging (8) gives $f(N, \mathcal{F}_0) = (1 - P_0)^{1/S}$. Using the dynamic programming solution to (7), we build a two-dimensional table of values of $f(N, \mathcal{F}_0)$ indexed by N and by a range of values of $\mathcal{F}_0$. For given N, S, and P_0, search row N of this table for the value of $\mathcal{F}_0$ such that $f(N, \mathcal{F}_0) \approx (1 - P_0)^{1/S}$. More accurate values of $\mathcal{F}_0$ can be found by interpolation. If there are nm entries in this table, where n and m are the number of values represented for N and F_0, respectively, then the table can be constructed off-line in time $O(nmN^3)$. Here, N represents the maximum number of data points. On-line, an $O(\log m)$ binary search finds the appropriate value of $\mathcal{F}_0$ for given values of P_0, S, and N.

Since we derived $\mathcal{F}_0$ from P_0, the desired probability the best fit to random data is less than $\mathcal{F}_0$, this threshold prevents MINPRAN from hallucinating. We may also use our analysis to demonstrate why MINPRAN can accept fits involving arbitrarily low inlier percentages. Given P_0, N, and S, which depends on P_g, we find $\mathcal{F}_0$ as described above, and for each i, $1 \le i \le N$, we find s_i such that $\mathcal{F}(s_i, i, N) = F_0$. As long as a fit with i inlier residuals ($i + p$ inliers total) has residual bound $r \le s_i$, it may be distinguished from a random fit and therefore accepted as correct. Table II gives example $\mathcal{F}_0$ and s_i values.

TABLE II
VALUES OF s_i (WITH $Z_0 = 1$) SUCH THAT $\mathcal{F}(s_i, i, N) = \mathcal{F}_0$ FOR $N = 50$, $P_0 = 0.05$, $Z_0 = 1.0$, AND $S = 25$ (SO THAT $\mathcal{F}_0 = 0.000095$) OR $S = 50$ (SO THAT $\mathcal{F}_0 = 0.000045$). WHEN $Z_0 \neq 1$ THESE s_i MUST BE SCALED BY Z_0

i	s_i	
	$S = 25$	$S = 50$
5	0.009	0.008
15	0.102	0.096
25	0.243	0.236
35	0.428	0.416
45	0.661	0.647

IV. ANALYSIS OF MINPRAN

The derivation of $\mathcal{F}_0$ in the previous section establishes MINPRAN's ability to distinguish between good fits to good data and fits to purely random data. In this section, we analyze MINPRAN's performance when a good fit should be found to show it should find good fits and nearly the correct number of inliers. To make this analysis tractable, we break it up into two steps, showing

1) for the correct fit, the expected minimum of $\mathcal{F}(r_{\phi,i}, i, N)$ occurs at nearly the correct number of inliers (see Section IV.A), and
2) when the residuals take on their expected values relative to any fit, the correct fit, ϕ, minimizes $\mathcal{H}(\phi, N)$ (see Section IV.B).

The former allows MINPRAN to generate the best possible final least-squares fits, including good estimates of the variance of the data. While neither of these models the performance of MINPRAN exactly, they are strongly indicative of its behavior. The first approximates MINPRAN's behavior when a sample containing all inliers to the correct fit is tested. The second compares criterion function values across all possible fits.

A. Finding (Nearly) the Correct Number of Inliers For the Correct Fit

We show for the correct fit ϕ when k inliers are expected, the minimum of $\mathcal{H}(\phi, N)$ occurs at almost exactly k inliers. To do this, we derive the probability that the minimum of the criterion function occurs at i, $1 \le i \le N$. Based on this, we compute the expected number of inliers, and the variance in the number of inliers. In the derivation, we assume

1) the k expected inlier residuals are Gaussian with zero mean and standard deviation σ, and
2) the $N - k$ expected outlier residuals are uniformly distributed in the range $[0 \ldots Z_0]$. (Observe that when k inliers are expected and the points are independent, the actual number of points taken from the inlier process follows a binomial distribution.)

Normalizing so that $Z_0 = 1.0$, let $p_{min}(i, r)$ describe the probability density function (pdf) of the criterion function minimum occurring at i inliers, $1 \le i \le N$, and inlier bound, r, $0 \le r \le 1$. Then the probability the minimum occurs at any given i is

$$p(i) = \int_0^1 p_{min}(i, r)dr \quad (9)$$

The expected number of inliers and the variance in the number of inliers are

$$E[i] = \sum_{i=1}^{N} i\, p(i) \quad \text{and} \quad Var[i] = \left\{\sum_{i=1}^{N} i^2 p(i)\right\} - (E[i])^2. \quad (10)$$

We derive $p_{min}(i, r)$ by generalizing the derivation of $f(N, \mathcal{F}_0)$ in Section III.B. For given values of i and r, let $\mathcal{F}^* = \mathcal{F}(r, i, N)$ and let $f_1, \ldots f_N$ be the unique values such that $\mathcal{F}(f_j, j, N) = \mathcal{F}^*$. Given N residuals $p_1, \ldots, p_N$ in increasing order, two different sets of constraints must be satisfied to produce a minimum of $\mathcal{F}^*$ at i, r. First, $p_j > f_j$, $1 \le j \le N$ and $j \neq i$, to force $\mathcal{F}(p_j, j, N) > \mathcal{F}^*$ for all j. Second, $p_i = r = f_i$ to force $\mathcal{F}(p_i, i, N) = \mathcal{F}^*$. These imply there must be $i - 1$ residuals in the range $[0 \ldots r)$, one residual at r, and $N - i$ residuals in the range $(r \ldots 1]$.

To derive the probability these constraints are satisfied, let $f_0 = 0$ and $f_{N+1} = 1$, and, as in Section III.B, define the interval counts, c_j, to be the number of residuals in the interval $(f_j \ldots f_{j+1}]$. The only exception is c_{i-1}, which we define as the

number of residuals in $(f_{i-1} \ldots f_i)$ to exclude the location $f_i = r$ from the interval counts. Based on these definitions, the first set of constraints becomes

$$\sum_{j=0}^{t} c_j \le t,\ 0 \le t < i-1, \text{ and } \sum_{j=0}^{i-1} c_j = i-1 \tag{11}$$

$$\sum_{j=i}^{t} c_j \le t-i,\ i \le t < N, \text{ and } \sum_{j=i}^{N} c_j = N-i. \tag{12}$$

These equations are similar to (6) in Section III.B, and are illustrated in Fig. 4. The probability of having c_j points in the interval $(f_j \ldots f_{j+1}]$ is governed by a pdf combining the expected k Gaussian inliers and the expected $N-k$ uniform outliers,

$$g(x) = \frac{k}{N}\frac{\sqrt{2}}{\sqrt{\pi}\sigma} e^{-x^2/2\sigma^2} + \frac{N-k}{N}. \tag{13}$$

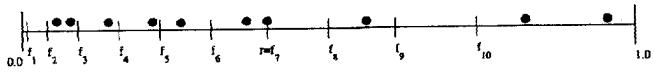


Fig. 4. A set of residuals that satisfies the constraints to have a minimum of $\mathcal{F}$ at i and r $(i = 7)$. (Each dot represents a residual value, p_j.) For each j, $p_j \ge f_j$, and $p_7 = f_7 = r$.

The range of x is assumed to be $[0 \ldots 1]$.[4] The inlier pdf and the outlier pdf are independent and defined over the same range because they are assumed to arise from different physical processes. The probability c_j particular (independent) points are in the interval $(f_j \ldots f_{j+1}]$ (or $(f_{i-1} \ldots f_i)$ for c_{i-1}) is simply

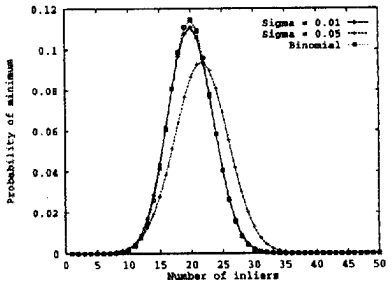
$$\left[\int_{f_j}^{f_{j+1}} g(u)du\right]^{c_j}. \tag{14}$$

Now we are ready to calculate $p_{min}(i, r)$. Let C_l be the set of all $c_0, \ldots, c_{i-1}$ satisfying (11) and let C_u be the set of all $c_i, \ldots, c_N$ satisfying (12). Then, by analogy with (7)

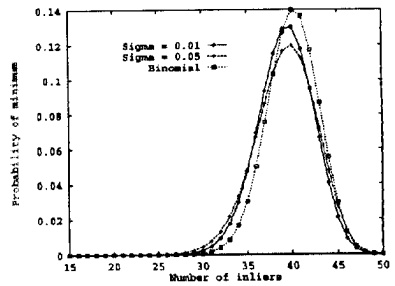
$$p_{min}(i, r) = \sum_{\substack{(c_0,\ldots,c_{i-1})\in C_l,\\ (c_i,\ldots,c_N)\in C_u}} \binom{N}{c_0, \ldots, c_{i-1}, 1, c_i, \ldots, c_N} g(r)\prod_{j=0}^{n}\left[\int_{f_j}^{f_{j+1}} g(u)du\right]^{c_j}, \tag{15}$$

where the 1 in the multinomial coefficient represents choosing the residual $f_i = r$, and $g(r)$ is the density of that residual. This equation may be evaluated in $O(N^3)$ time using a dynamic programming technique [31].

As discussed above, we use the equation for $p_{min}(i, r)$ to calculate $p(i)$, $E[i]$, and $Var[i]$. Plots of $p(i)$ are shown in Fig. 5 for $k = 20$ and $k = 40$ expected inliers, when there are 50 data points, and $\sigma = 0.01$ and $\sigma = 0.05$. A third curve in each plot shows the binomial probability of i actual inliers when k are expected, representing the ideal model for $p(i)$. For the two values of σ when $k = 20$, $E[i]$ is 19.95 and 21.58, respectively, and the standard deviation $(Var[i])^{1/2}$ is 3.58 and 4.30 points. The peak in the curves shifts to the right as σ increases because the inlier bound is larger and more outliers are likely to fall inside the bound. For the two values of σ when $k = 40$, $E[i]$ is 39.3 for both, and $(Var[i])^{1/2}$ is 3.04 and 3.35. These values predict a slight bias toward fewer inliers found by MINPRAN when there is a large percentage of expected inliers. Overall, the results show that when testing the correct fit, MINPRAN should find nearly the correct number of inliers for a wide range of values of σ and expected number of inliers. These results are confirmed by our experimental results which show a close match between the expected number of inliers predicted from (10) and the actual number of inliers found by MINPRAN.



a



b

Fig. 5. Two sets of plots of $p(i)$ (9) for $N = 50$ points, and $\sigma = 0.01, 0.05$. Plot (a) shows $k = 20$ expected inliers, while plot (b) shows $k = 40$ expected inliers. The third plot shows the binomial probability that i is the actual number of inliers when $k = 20$ inliers are expected.

B. The Minimum of $\mathcal{F}$ Based on the Expected Residuals

To predict MINPRAN's behavior when it compares fits to minimize $\mathcal{H}(\phi, N)$, our second analysis technique derives the expected residuals to any given fit, evaluates $\mathcal{H}(\phi, N)$ for these expected residuals, and searches for the minimum of $\mathcal{H}$ across all possible fits. We have applied this same technique to evaluating a variety of robust estimators used in computer vision. Details of the technique and results are given in [28], so we only present a summary here.

Given a model of the data in two-dimensions, we derive a density $h(x, z)$ of the data points in $\Re^2$ and use this to calculate the expected residuals relative to each fit. Density $h(x, z)$ combines models of the inlier process(es) of the curve(s) generating the data (including models of the density of points on each curve and the noise in each point's z value) and the outlier process. For any hypothesized fit $\phi(x)$ and residual r, we calculate $F(r \mid \phi)$, the cumulative distribution of r relative to ϕ,

4. Strictly speaking, placing an upper bound of 1 on x violates the assumption of having Gaussian inliers since the Gaussian is unbounded. Practically, however, this makes little difference since σ is presumably small relative to Z_0.

by integrating $h(x, z)$ over the region $\phi(x) \pm r$; we calculate $f(r \mid \phi)$, the density of r relative to ϕ, as the derivative of $F(r \mid \phi)$. We then derive the conditional expected absolute residuals, $e_{\phi,i}$, $i \le i \le N$, from $F(r \mid \phi)$ and $f(r \mid \phi)$ by applying standard order statistics techniques [7].

For each possible fit ϕ, we define $AEV(\phi, N) = \min_i \mathcal{F}(e_{\phi,i}, i, N)$. This is the "approximate expected value" of $\mathcal{H}(\phi, N)$, i.e. the value of the randomness criterion when the residuals take on their expected values relative to ϕ. Finally, we search (numerically) over the range of possible fits to find $\min_\phi AEV(\phi, N)$. The fit producing the minimum approximates the expected best fit to data generated from the curve and outlier models.

Using this technique, we obtain both good and bad results. When a single curve, β, generates the data and the remaining points are random outliers, the AEV is minimized when $\phi = \beta$. Intuitively, this occurs simply because the correct fit produces the smallest expected residuals, and therefore, the smallest criterion values. Combined with our analysis in Section IV.A, this shows that when a single curve generates the good data MINPRAN should find the correct fit, including approximately the correct set of inliers. With this inlier set, MINPRAN's final least-squares fit should produce accurate fit parameters and good estimates of the noise in the data.

Unfortunately, when multiple curves generate the data, for example near a depth or orientation discontinuity in a range image, the minimum AEV usually occurs at a fit that "bridges" the discontinuity to intersect both curves, and whose inlier bound expands to include nearly all the points from both curves. These "bridging fit" errors, which are characteristic of many existing robust estimators used in vision [28], occur in MINPRAN for a wide range of discontinuity magnitudes [29].

In Section VII, where we discuss applying MINPRAN to range data, we introduce two simple techniques that avoid most bridging fit errors. First, however, we describe simulation results confirming our analysis of MINPRAN and comparing it to LMS, and then, as a preliminary to applying MINPRAN, we generalize its outlier model to nonuniform outlier distributions.

V. Simulation Results

In order to confirm several of MINPRAN's properties—it avoids hallucinating, it finds good fits for a wide range of expected inlier percentages, and it estimates more accurate fits than least median of squares (LMS)—we analyze its ability to fit planar surfaces, $z = a_0 + a_1x + a_2y$, to synthetic range data. See Fig. 6 for a preliminary result.

In the first set of experiments all of the data points are uniformly distributed outliers. Using $P_0 = 0.1$ (corresponding to a desired 10% probability of accepting a fit to random data; see Section III.B), MINPRAN accepted a random fit as correct (hallucinated) in only 1.6% of 2500 test data sets. (These results show the independence assumption in Section III.B is conservative since it leads to overestimating the probability that the minimum value of $\mathcal{H}$ is less than $\mathcal{F}_0$, and therefore leads to a lower value of $\mathcal{F}_0$.) Other robust techniques, including LMS, always find a single fit, even when the data are purely random.

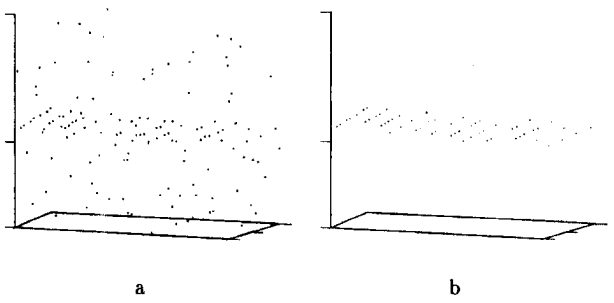


Fig. 6. Example experimental result on the image shown in Fig. 1a. Plot (a) shows the original data and plot (b) shows the reconstructed set of points MINPRAN labeled as inliers. Of the 144 data points, 58 were generated from a planar surface; MINPRAN's best fit included 56 inliers.

In the second set of experiments, a planar surface generates points, with expected inlier percentages of 20%, 25%, ..., 90%; the remaining points being outliers. In each test, 100 data points are generated, and the sampling parameters are $P_g = 0.99$, $n_f = 1$, and $x_0 = p_e - 0.1$, where p_e is the expected fraction of outliers. The depth and noise parameters are $Z_0 = 100$ and $\sigma = 1.0$. 2,500 tests were run for each inlier percentage. The results are summarized in Fig. 7.

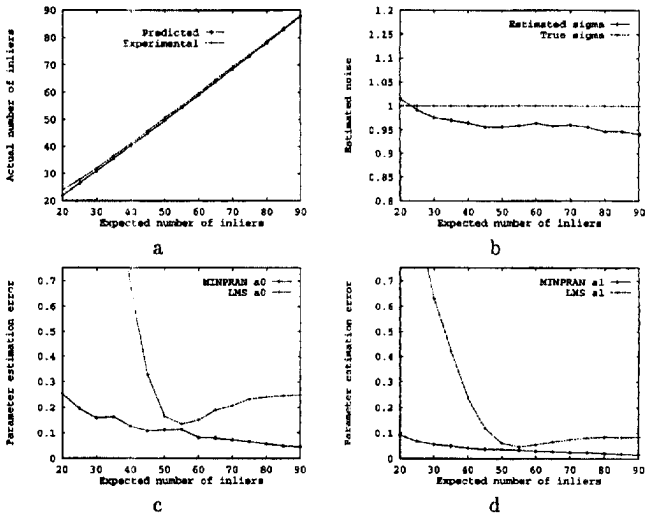


Fig. 7. Summary of experimental results in finding a single fit when $k = 20, 25, \ldots, 90$ inliers are expected out of 100 points. Plot (a) shows the average number of inliers found by MINPRAN along with the number of inliers predicted using the techniques of Section IV.A. Plot (b) shows the average estimated value of the standard deviation of the noise in the data. Plots (c) and (d) show the average parameter errors for MINPRAN and LMS relative to the least sum of squares fits to the known inliers for the data sets.

Fig. 7a shows that the average number of inliers closely matches predictions made using the technique of Section IV.A. The average is slightly higher than predicted for small expected inlier percentages because when the actual data contains fewer than the expected number of inliers, there is a greater chance the fit will be missed, either during the random sampling process or by failing the randomness threshold. This results in a bias toward more inliers in the fits accepted as correct. As evidence for this, acceptable fits were found in 88% of the data sets for $k = 20$, in 94% for $k = 25$, and in 100% for $k = 30$ and higher.

Fig. 7b shows the average estimated value of σ, the standard deviation of the noise in the data. MINPRAN underestimates σ by about 5%, on average. This is because, using our outlier model, some of MINPRAN's inliers may be true outliers closer to the fit than the good points having the largest residuals, and because, conversely, a small number good points having the largest residuals may be outside the inlier bound. After MINPRAN's least-squares fit, we correct for missing the good points with the largest residuals by gathering all points within $\pm 3\sigma$ of the fit. Applying a second least-squares fit to all gathered points corrects for underestimating σ.

Figs. 7c and 7d plot the average errors in the fit parameters for MINPRAN and LMS as a function of the expected number of inliers. Plots are shown for the intercept parameter, a_0, and one of the slope parameters, a_1. The errors are the absolute differences between the estimated parameters and the parameters of the least sum of squares fit to the known inliers. These results show MINPRAN finds good fits for a wide range of inliers, and MINPRAN finds substantially better fits than LMS.

VI. Nonuniform Outlier Distributions

We derived MINPRAN's criterion function, $\mathcal{H}$, by assuming the outliers are drawn from a uniform distribution over the depth range. In practice, however, sensor geometry may introduce bias in the outliers' depth values. For example, triangulation-based sensors, such as stereo cameras or structured-light range finders, measure the disparity, d, between point positions in multiple "viewpoints," and convert this disparity to depth, z, using a calibration model. Often, $z \propto 1/d$. Empirically, we have found the outliers in disparity to be approximately uniformly distributed, implying the density of outliers in z is proportional to $1/z^2$ (using standard techniques to transform densities [18]). This violation of our uniform outlier model, which arises when we triangulate to convert the measured values, d, to the desired values, z, can be handled in two ways. The first is to reconstruct in disparity rather than in depth. This is reasonable because the $1/z$ transform biases the noise in the data as well as the outliers. The second is to modify the criterion function to handle nonuniform outlier distributions. We show how to do the second now.

Looking back at (1) which defines $\mathcal{F}(r, k, N)$ and forms the basis of $\mathcal{H}(\phi, N)$, the value r/Z_0 in the binomial distribution is the probability a random outlier could fall within the inlier band around a fit ϕ. To handle nonuniform outlier distributions, we replace r/Z_0 by the probability, $P_z(r \mid \phi)$, that an outlier could fall within the region defined by $\phi \pm r$. We refer to $P_z(r \mid \phi)$ as the "residual probability." $P_z(r \mid \phi)$ is the integral of the outlier density over the region bounded by $\phi \pm r$ (the domain of the data and the depth range). In general, when the outlier density varies across the region, this integral can be difficult to evaluate.

One approximation to $P_z(\phi, r)$, based on assuming the outlier density varies slowly within $\phi \pm r$, is to integrate the outlier density at the data domain's center. Specifically, for data taken from an image, let (x_c, y_c) be the data's center of mass, let $z_c = \phi(x_c, y_c)$, and let $p_z(z \mid x_c, y_c)$ be the outlier density at (x_c, y_c). Then, we approximate $P_z(\phi, r)$ by

$$\int_{z_c - r}^{z_c + r} p_z(z \mid x_c, y_c) dz.$$

As an example, when $p_z(z \mid x_c, y_c) = a/z^2$ (a is a normalizing constant), the approximation yields

$$p_z(r \mid \phi) = \frac{2ar}{z_c^2 - r^2}.$$

In this example, calculating the residual probabilities is only slightly more expensive than calculating the residuals themselves.

The residual probability $P_z(r \mid \phi)$ is the cumulative distribution of r given ϕ (based on the outlier density); hence, it is itself uniformly distributed ([24], p. 141). Thus, after replacing residuals by residual probabilities in (1), the rest of MINPRAN is unchanged. For each fit, MINPRAN calculates and sorts the residual probabilities for each of the N fit residuals (for a single fit this ordering is actually the same as the ordering of the residuals); it then saves, for each i, the fit with the smallest residual probability; and finally, it evaluates $\mathcal{F}$ for each of these N retained residual probabilities, choosing the fit associated with the minimum as the best fit. In deciding whether or not the best fit is random, MINPRAN uses the same threshold, $\mathcal{F}_0$, derived for the uniform distribution, since in $\mathcal{F}_0$'s derivation the residuals are essentially residual probabilities.

We have, therefore, generalized MINPRAN to arbitrary outlier distributions. Except for calculating $P_z(\phi, r)$, which may need to be approximated to compute it efficiently, the generalization is straightforward.

VII. Application to Range Data: MINPRAN2

Having defined MINPRAN and studied its properties both theoretically and using simulated data, we now apply it to range data taken from complicated scenes, fitting surface patches, eliminating outliers, and smoothing inliers. As demonstrated in Figs. 10 and 11, these data may contain multiple overlapping surfaces, numerous discontinuities which lead to frequent outliers, and regions, which due to occlusions or illumination conditions, contain only outliers. Thus, these data exhibit the complexities demonstrated in Section I.

We assume the data form a range image, with each pixel containing either a depth measurement or a special value indicating no measurement is available. We apply MINPRAN to each of a series of regularly spaced, overlapping image regions, by first extracting a set of points from the depth measurements in each region, and then applying the algorithm to each point set. The regions overlap because large regions are required to obtain reasonably accurate fits and the resulting fits decrease in accuracy toward region boundaries [17].

A. The Bridging Fits Problem

To apply MINPRAN successfully, we need to address the bridging fits problem, MINPRAN's tendency to bridge discontinuities and include points from multiple surfaces as inliers to its best fit. We do so using two complementary solutions:

1) Within each individual region, to avoid selecting a bridging fit, the "split-search" technique finds both the single best fit to the data and two good disjoint fits. It then uses a modified randomness criterion to choose between the single fit and the pair of fits, generally preferring the pair of fits only when the data arise from multiple surfaces.
2) To eliminate bridging fits after applying MINPRAN to all regions, the "final randomness test" identifies the best fit to each point, removes each point from the inlier set of any fit that is inconsistent with its best fit, and reevaluates the randomness of each fit based on its remaining inliers. It eliminates any fits that fail their original randomness threshold, $\mathcal{F}_0$.

The final randomness test successfully eliminates a bridging fit when there are correct fits in neighboring regions. It therefore eliminates most bridging fits at isolated depth and orientation discontinuities, failing on depth discontinuities only when the discontinuity magnitude is less than 4σ [32]. It can not, however, eliminate bridging fits arising from overlapping surfaces (see Fig. 1c) that extend across multiple regions (such as seen in Fig. 10). Fortunately, when the overlapping surfaces are separated in depth by at least 5σ, split-search generally estimates two correct fits, one for each surface. Split-search, however, is less sensitive to isolated discontinuities than the final randomness test, tending to choose bridging fits when the step height is less than 8σ. (See [32] for complete analysis.) Thus, split-search and the final randomness test effectively complement each other, and together they form the final version of MINPRAN, which we call MINPRAN2. We sketch the details of these techniques below.

A.1. Split-Search

For the split-search technique, we first describe the modified criterion function. In a given region, let ϕ_b be the fit minimizing $\mathcal{H}$, with k_b inliers and an inlier bound of r_b, and let ϕ_1 and ϕ_2 be a pair of disjoint fits, with k_1 and k_2 inliers and residual bounds r_1 and r_2. (See Fig. 8 for a two-dimensional example.) Since $\mathcal{H}$ is minimized by ϕ_b rather than either ϕ_1 or ϕ_2 individually, we want to compare ϕ_1 and ϕ_2 together to ϕ_b.

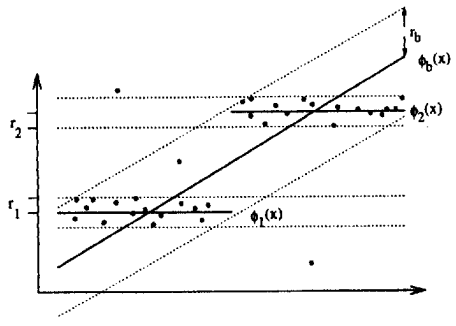


Fig. 8. The inlier bounds for the bridging fit, $\phi_b(x)$, and the two correct fits for a step edge, $\phi_1(x)$ and $\phi_2(x)$. The dotted lines show the inlier bounds.

The probability at least k_1 points could randomly be within $\phi_1 \pm r_1$ and at least k_2 points could randomly be within $\phi_2 \pm r_2$ is

$$\sum_{i=k_1+k_2}^{N} \sum_{j=k_2}^{i-k_1} \frac{N!}{(i-j)!\,j!\,(N-i!)} \left(\frac{r_1}{Z_0}\right)^{i-j} \left(\frac{r_2}{Z_0}\right)^{j} \left(1-\frac{r_1+r_2}{Z_0}\right)^{N-i}, \quad (16)$$

where i represents the number of points in both regions, j represents the points in $\phi_2 \pm r_2$, and $i-j$ represents the points in $\phi_1 \pm r_1$. It is easy to show this is less than

$$\sum_{i=k_1+k_2}^{N} \binom{N}{i} \left(\frac{r_1+r_2}{Z_0}\right)^{i} \left(1-\frac{r_1+r_2}{Z_0}\right)^{N-i} = \mathcal{F}(r_1+r_2,\, k_1+k_2,\, N), \quad (17)$$

which is the probability there are at least $k_1 + k_2$ points in the combined regions.[5] Then, based on the probability of randomness, we would like to have MINPRAN choose the two correct fits instead of the bridging fit if

$$\mathcal{F}(r_1 + r_2, k_1 + k_2, N) < \mathcal{F}(r_b, k_b, N). \quad (18)$$

Equation 18 allows us to compare a single fit to a pair of disjoint fits. To actually find the disjoint fits to make this comparison, we must be careful to avoid a combinatorial explosion in the search. Our heuristic technique uses the following simple observation: if there are two surfaces in a region, then at least one of them contains less than half of the points. Based on this, we modify MINPRAN to save two fits from its examination of S fits: the single best fit as before, and the best fit containing fewer than half of the data. MINPRAN then temporarily assumes the latter is correct, gathers and marks its inliers, and repeats the search (as described in Section III), finding the single best fit in the remaining data. (For example, when one half of the step edge is found initially, the second half will be found in the second search.) These two fits may then be compared to the initial best fit using (18) . This technique adds relatively little to the computation time since the second search involves significantly fewer points and samples.

A.2. The Final Randomness Test

The final randomness test depends on the overlap between image regions in which MINPRAN estimates fits, and is based on several observations. First, when a bridging fit is found in one region, its inliers may also be part of correct fits in overlapping regions, as shown in Fig. 9. Next, because a bridging fit usually incorporates *all* inliers within its region from both surfaces, it produces erroneous fit parameters and overestimates the variance in the data. The fit parameter errors shift the estimated z values of the inliers away from their correct values. Finally, a point should not be considered an inlier to two fits from different regions if they produce significantly different estimates of its final z value. In this case, the fit producing the minimum variance estimate should determine the z value, and the point should not be an inlier to the other fit.

5. Depending on the values of k_1, r_1, k_2, and r_2, $F(r_1 + r_2, k_1 + k_2, N)$ may be a gross overestimate of (16). Since we will force nonrandomness conditions on the individual fits, however, in practice the difference between the two is relatively insignificant.

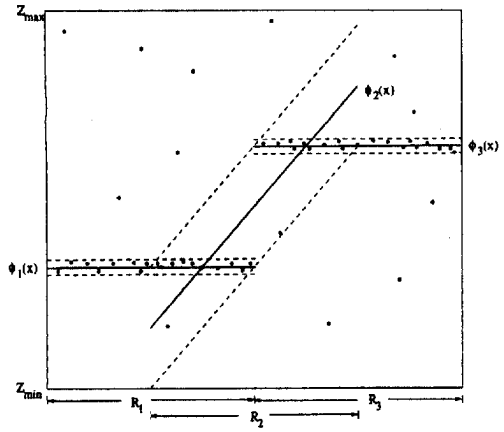


Fig. 9. Example of a bridging fit, ϕ_2, for a step edge in one region, R_2, and correct fits, ϕ_1 and ϕ_3, in overlapping regions, R_1 and R_3. The dashed lines show the inlier bounds.

Based on these observations, we say a point p_i is a "true inlier" to fit ϕ if ϕ produced the minimum variance z estimate at p_i or if ϕ is "consistent" with the fit that produced the minimum variance estimate. Consistency depends on both

1) ϕ's z estimate being within the uncertainty range of the best estimate, and
2) ϕ's estimate of the variance in the data being within the χ^2 uncertainty range of the best estimate.

When there are correct fits in overlapping regions, bridging fits will keep few true inliers.

After determining the true inliers to each fit, the final randomness test re-evaluates the criterion function for each fit based only on its true inliers. Suppose fit ϕ has k true inliers, suppose the largest absolute residual of these k points relative to ϕ is r, and let $\mathcal{F}_0$ be the original randomness threshold in ϕ's image region. Then, if $\mathcal{F}(r, k, N) > \mathcal{F}_0$, fit ϕ is eliminated.[6]

B. Experimental Results

In its present implementation, MINPRAN2 fits planar patches of the form $z = a_0 + a_1x + a_2y$ to range data acquired using a structured light range finder [22]. It outputs both the reconstructed data points and the surface patches. Each data point considered a true inlier to at least one surviving fit is output, with its final z value taken as the weighted average of the z estimates made by the fits to which it is considered an inlier. The weight is inversely proportional to the fit's uncertainty at the point. For each surface, both the fit parameters and a rectangle bounding the image coordinates of its true inliers are output.

Two representative results on range images of complicated scenes are shown in Figs. 10 and 11. Each figure shows an intensity image of the scene, color images of the range data before and after reconstruction, and three-dimensional views of the data, the reconstructed data, and the surface patches. (The range data and results are shown as disparity values, emphasizing the uniformity of disparity outliers. We also reconstruct using disparities to demonstrate the first method of handling nonuniform outlier distributions discussed in Section VI.) The bright rectangular regions in the center of each scene intensity image show the illumination window of the sensor. Outside this region the data are entirely random. We left these regions uncropped to demonstrate MINPRAN's ability to ignore purely random data. (Due to sensor geometry, the depth range shifts from the top to the bottom of the range images, moving from nearer (red) at the top to farther at the bottom; hence the shift in the colored depth values from red at the top to blue at the bottom.) For the results presented, MINPRAN2 fit planar patches in 10×10 image regions spaced by 5 pixels in each dimension. It required at most 40 seconds on a Sparc-10 for each of these images.

The most challenging part of Fig. 10 is the area covered by the racquetball racket's strings. Through these strings, points from the background, the baseball glove, and a racquetball all appear, as well as several outliers (see the red points in the three-dimensional view of the string area). Using the split-search technique, MINPRAN2 was able to fit multiple surfaces in the regions within this area, with one fit including points from the strings, and at least one other fit including points from the partially occluded surface. Although they do not appear in the plot of the surface patches, MINPRAN2 fit patches to points from two different racquetballs. Near where the racket leaned on the glove, one or two fits bridging points from the racket strings and the glove survived the final randomness test. MINPRAN2 eliminated nearly all outliers in the area covered by the strings. Finally, several of the surface orientations appear to be inaccurate, either due to the inclusion of one or two outliers, or due to having a small number of inliers.

The wires range image in Fig. 11 challenges MINPRAN2's ability to ignore regions where there are only bad data. The wires near the bottom of the bright region in the scene intensity image are actually outside the illumination range of the sensor, even though they are in the camera's projection of the illuminated region. Thus, only bad data appear in these regions. As is obvious from both the color images and the 3D views of the reconstructed data, MINPRAN2 identified all these points as outliers and eliminated them. Most of the wire data was cleanly reconstructed, although a few isolated errors remained and some sections of the wires containing too few points were missed. In the figure, the surface patches occasionally appear cluttered, a few orientations are inaccurate, and, on the bottom right of the image, there is an outright error. The clutter is mostly due to the overlap of the wires and the power box behind the wires. As with the sports equipment, fits with inaccurate orientations either have only a few inliers or include one or two outliers. Finally, the erroneous fit is a surface bridging the foreground and a small number of distant outliers. MINPRAN2 tolerates this error because it has no notion of the distribution of points along a surface.

To summarize, these results (and others not shown) demonstrate MINPRAN2's ability to effectively fit surface patches and reconstruct range data taken from complicated scenes. MINPRAN2 eliminated most outliers, avoided hallucinating fits, and reconstructed multiple overlapping fits where neces sary. On the negative side, it occasionally obtained fits that included points from multiple surfaces or included points from a surface and one or two outliers.

6. We use $\mathcal{F}$ instead of $\mathcal{H}$, which is the minimum of $\mathcal{F}$ over all numbers of inliers and inlier bounds, because the number of inliers is fixed at k.

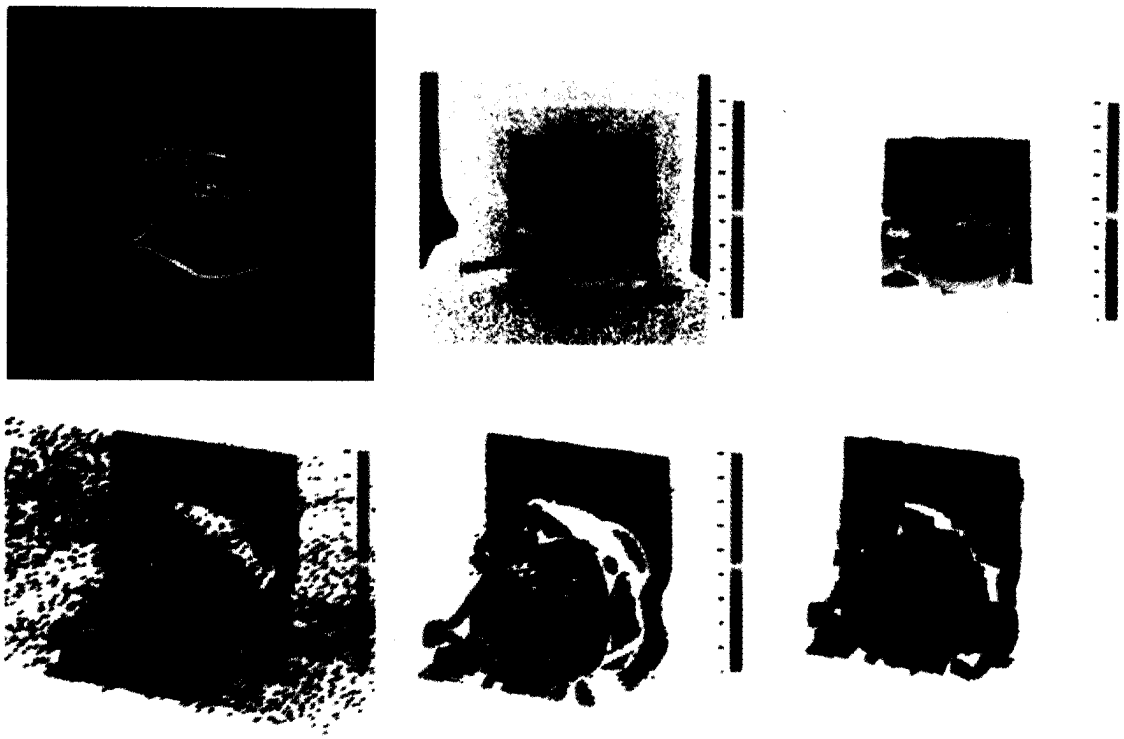


Fig. 10. Results for the "sports equipment." Top left: gray level image showing the illumination area. Top center: color image of the range (disparity) data. Top right: color image of the reconstructed inliers. Bottom left: colored, three-dimensional view of the original data. Bottom center: colored, three-dimensional view fo the reconstructed inliers. Bottom right: three-dimensional shaded view of the surface patches.

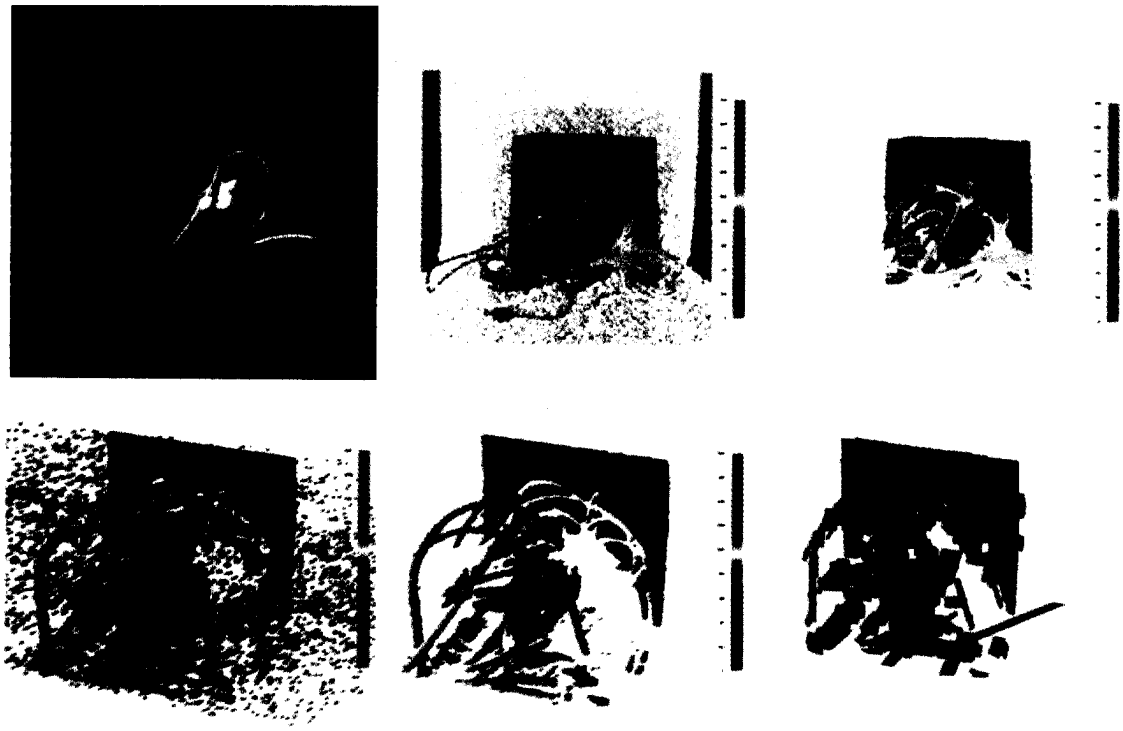


Fig. 11. Results for the "wires." Top left: gray level image showing the illumination area. Top center: color image of the range data. Top right: color image of the reconstructed inliers. Bottom left: three-dimensional view of the original data. Bottom center: three-dimensional view of the reconstructed inliers. Bottom right: three-dimensional shaded view of the surface patches.

VIII. Summary and Discussion

In this paper we have derived a new robust estimator, MINPRAN, analyzed its properties both theoretically and experimentally, and applied it to reconstructing surfaces and eliminating outliers in complicated range data. MINPRAN is the first technique that reliably tolerates more than 50% outliers without assuming a known inlier bound; it only assumes the outliers are uniformly distributed within the dynamic range of the sensor. Since assumptions about the outlier data are required to surpass the 50% breakdown point, the uniform distribution assumption is, in a sense, the minimal possible assumption. Further, when the outliers are nonuniform—for example when there are biases in the depth values produced by a range sensor, we have shown how to adjust MINPRAN's randomness criterion function accordingly.

MINPRAN has several novel features and compares favorably to least median of squares (LMS), the robust estimator that achieves the highest breakdown point of the estimators that do not assume a known noise model for the good data. MINPRAN generally does not "hallucinate" fits when there are none in the data, and it estimates good fits, including approximately the correct number of inliers, for both low and high inlier percentages. By contrast, LMS always estimates a single fit involving 50% of the data points, regardless of the true outlier percentage. The cost of MINPRAN's flexibility is a moderate increase in its computational complexity relative to LMS, ($O(SN \log N + N^2)$ vs. $O(SN \log N)$).

In applying MINPRAN to range data, we introduced both the split-search technique and final randomness test to avoid making bridging fit errors. The resulting estimator, MINPRAN2, reliably eliminates outliers, does not hallucinate, and estimates multiple surface patches where appropriate. Unfortunately, it also occasionally retains surfaces bridging small magnitude discontinuities.

Thus, although the main contribution of this paper is the theoretical development of MINPRAN, MINPRAN2 may be used in practical applications. While it only estimates local surface patches, meaning it is not a complete reconstruction technique, it tolerates higher outlier percentages than other techniques and it can estimate fits to data from multiple, overlapping surfaces. MINPRAN2 could therefore be used to seed surfaces in surface growing algorithms [4], [5], [6], [13], or it could be used to "clean" the data for global reconstruction algorithms, especially recent ones that include overlapping surfaces in their optimization model [14]. These possibilities are being explored as part of ongoing work.

Acknowledgments

The author would like to acknowledge the financial support of the National Science Foundation under grant IRI-9217195, and the assistance of Robin Flatland and James Miller for their detailed comments on earlier drafts of this paper.

References

[1] M. Abramowitz and I.A. Stegum, *Handbook of Mathematical Functions*. Dover Publications, 1968.
[2] D.H. Ballard and C.M. Brown, *Computer Vision*. Prentice Hall, 1982.
[3] P.J. Besl, J.B. Birch, and L.T. Watson, "Robust window operators," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 591-600, 1988.
[4] P.J. Besl and R.C. Jain, "Segmentation through variable-order surface fitting," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, pp. 167-192, 1988.
[5] T. Boult and M. Lerner, "Energy-based segmentation of very sparse range surfaces," *Proc. IEEE Int'l Conf. Robotics and Automation*, pp. 232-237. 1990
[6] K.L. Boyer, M.J. Mirza, and G. Ganguly, "The robust sequential estimator: A general approach and its application to surface organization in range data," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, pp. 987-1,001, 1994.
[7] H.A. David, *Order Statistics*. John Wiley & Sons, 1970.
[8] M.A. Fischler and R.C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *CACM*, vol. 24, pp. 381-395, 1981.
[9] W.E.L. Grimson and D. Huttenlocher, "On the sensitivity of the Hough transform for object recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, pp. 255-274, 1990.
[10] J. Illingworth and J. Kittler, "A survey of the Hough transform," *CVGIP*, vol. 44, pp. 87-116, 1988.
[11] J.-M. Jolion. P. Meer, and S. Bataouche, "Robust clustering with applications in computer vision," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, pp. 791-802, 1991.
[12] R. Kumat and A.R. Hanson, "Robust methods for estimating pose and a sensitivity analysis," *CVGIP: Image Understanding*, 1994.
[13] A. Leonardis, A. Gupta, and R. Bajcsy, "Segmentation as the search for the best description of the image in terms of primitives," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 121-125, 1990.
[14] S. Madarasmi, D. Kersten, and T.C. Pong, "Multilayer surface segmentation using energy minimization," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 774-775, 1993.
[15] P. Meer, D. Mintz, A. Rosenfeld, and D.Y. Kim, "Robust regression methods for computer vision: A review," *Int'l J. Computer Vision*, vol. 6, pp. 59-70, 1991.
[16] D. Mintz, "Robustness by consensus," Technical Rep. CAR-TR-576, Univ. of Maryland—Center for Automation Research, 1991.
[17] D.F. Morrison, *Applied Linear Statistical Methods*. Prentice Hall, 1983.
[18] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1984.
[19] G. Roth and M.D. Levine, "Extracting geometric primitives," *CVGIP: Image Understanding*, vol. 58, pp. 1-22, 1993.
[20] P.J. Rousseeuw, "Least median of squares regression," *J. American Statistical Assoc.*, vol. 79, pp. 871-880, 1984.
[21] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*. John Wiley & Sons, 1987.
[22] K. Sato and S. Inokuchi, "Range-imaging system utilizing nematic liquid crystal mask," *Proc. IEEE Int'l Conf. Computer Vision*, pp. 657-661, 1987.
[23] B.G. Schunck, "Robust computational vision," *Proc. Int'l Workshop Robust Computer Vision*, pp. 1-18, 1990.
[24] S.D. Silvey, *Statistical Inference*. John Wiley & Sons, 1975.
[25] S.S. Sinha and B.G. Schunck, "A two-stage algorithm for discontinuity-perserving surface reconstruction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, pp. 36-55, 1992.
[26] D.L. Souvaine and J.M. Steele, "Time- and space-efficient algorithms for least median of squares regression," *J. American Statistical Assoc.*, vol. 82, pp. 794-801, 1987.
[27] A. Stein and M. Werman, "Robust statistics in shape fitting," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 540-546, 1992.
[28] C.V. Stewart, "Expected performance of robust estimators near discontinuities," Tech. Rep. 94-10, Dept. of Computer Science, Rensselaer Polytechnic Inst., 1994.

[29] C.V. Stewart, "A new robust operator for computer vision: Application to range images," *Proc. IEEE Conf. Computer Vision and Pattern Recognition* pp. 167-173, 1994.

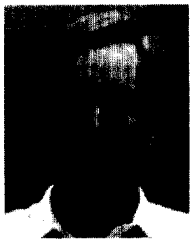
[30] C.V. Stewart, "A new robust operator for computer vision: Theoretical analysis," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1-8, 1994.

[31] C.V. Stewart, "A new robust operator for computer vision: Theoretical analysis," Technical Rep. 93-21, Dept. of Computer Science, Rensselaer Polytechnic Inst., Aug. 1993.

[32] C.V. Stewart, "A new robust operator for computer vision: Application to range and intensity images," Technical Rep. 93-24, Dept. of Computer Science, Rensselaer Polytechnic Inst., Oct. 1993.

[33] X. Zhuang and R.M. Haralick, "Developing robust techniques for computer vision," *Proc. Int'l Workshop Robust Computer Vision*, pp. 19-38, 1990.

[34] X. Zhuang, T. Wang, and P. Zhang, "A highly robust estimator through partially likelihood function modeling and its application in computer vision," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, pp. 19-35, 1992.



Charles V. Stewart (M'88) received the BA degree in mathematical sciences from Williams College in 1982 and the MS and PhD degrees in computer science from the University of Wisconsin in 1985 and 1988, respectively. He is currently an associate professor in the Department of Computer Science at Rensselaer Polytechnic Institute, Troy, New York. He is a member of the IEEE and the ACM. His research interests include computer vision, robust statistics, and computational geometry.