

# Expected Performance of Robust Estimators Near Discontinuities

Charles V. Stewart\*  
Department of Computer Science  
Rensselaer Polytechnic Institute  
Troy, New York 12180-3590  
stewart@cs.rpi.edu

## Abstract

*In extracting a polynomial surface patch near an intensity or range discontinuity, a robust estimator must tolerate not only the truly random bad data ("random outliers"), but also the coherently structured points ("pseudo outliers") that belong to a different surface. To characterize the performance of least median of squares, M-estimators, Hough transforms, RANSAC, and MINPRAN on data containing both random and pseudo outliers, we develop two analytical measures, "pseudo outlier bias" and "pseudo outlier breakdown". Using these measures, we find that each robust estimator has surprisingly poor performance, even under the best possible circumstances, implying that present estimators should be used with care and new estimators should be developed.*

## 1 Introduction

In computer vision applications, robust estimators must tolerate both random outliers and pseudo outliers when fitting a polynomial surface patch in the neighborhood of a range or intensity discontinuity. Random outliers are bad measurements, which may arise from specularities, boundary effects, physical imperfections in sensors, or low-level vision computation errors. Pseudo outliers are points belonging to a second surface, the surface on the other side of the discontinuity. They differ from random outliers because they exhibit coherent structure. (As shown in Figure 1, both random and pseudo outliers corrupt non-robust estimators, such as least-squares.) Although robust estimator performance is well-understood on data containing random outliers, their performance on data containing both random and pseudo outliers is not, and some preliminary experimental evidence

shows disappointing results [12]. Because discontinuities in vision data are common, we must understand robust estimator performance on data containing both types of outliers to use them effectively.

We introduce two new analytical performance measures, "pseudo outlier bias" and "pseudo outlier breakdown," to characterize robust estimators' ability to extract surface patches near discontinuities. Working from models of the inliers (the points belonging to the desired surface), the pseudo outliers, and the random outliers, these measures calculate the data's expected, ordered, absolute residuals relative to a given fit, use these residuals to calculate a representative value of the estimator's optimization criterion for that fit (approximating the criterion's expected value for that fit), and then search for the fit,  $\hat{\theta}$ , with the optimum representative value. We define pseudo outlier bias as the distance between  $\hat{\theta}$  and the expected fit to the inliers alone. If  $\hat{\theta}$  "bridges" the inliers and pseudo outliers, thereby giving essentially the same result as ordinary least-squares (Figure 1b), we say that pseudo outlier breakdown occurs. (See [16] for the relationship between pseudo outlier breakdown and bias and the analysis techniques from robust statistics.)

We use pseudo outlier breakdown and pseudo outlier bias to study the performance of M-estimators [9, Chapter 7], least median of squares (LMS) [11, 15], Hough transforms [10], RANSAC [4], and MINPRAN [18, 17] in extracting polynomial surface patches from range data. We study three idealized pseudo outlier models: step edges, crease edges and parallel surfaces. Step edges model depth discontinuities, where points from the upper half of the step are pseudo outliers to the lower half. Crease edges model surface orientation discontinuities, where points from one side of the crease are pseudo outliers to the other. Finally, parallel surfaces model transparent or semi-transparent surfaces, where a background surface appears through breaks in the foreground surface, and data from the background are pseudo outliers to the foreground.

\*This work was supported by NSF Grant IRI-9217195. The author would like to thank Kishore Bubna, Robin Flatland and James Miller for their assistance in preparing this paper.

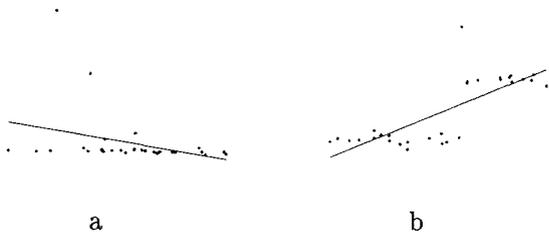


Figure 1: Examples demonstrating the effects of (a) random outliers and (b) pseudo outliers on linear least-squares fits.

## 2 Robust Techniques

Restricting our attention to regression models, we begin by defining the robust estimators we analyze. Our analysis techniques, and therefore our presentation, emphasize robust estimator optimization criteria and not search techniques. In describing the estimators, we assume the data are  $(\vec{x}_i, z_i)$ , where  $\vec{x}_i$  is a vector of image coordinates and  $z_i$  is a range value, and the fit is a function  $\theta(\vec{x})$ .

### 2.1 M-Estimators / Fixed-Band Methods

The optimization criteria of M-estimators of regression [9, Chapter 7] is

$$\min_{\theta} \sum_i \rho((z_i - \theta(\vec{x}_i))/\hat{\sigma}), \quad (1)$$

where  $\hat{\sigma}$  is an estimate of the true scale (noise) parameter,  $\sigma$ , and  $\rho(u)$  is a robust “loss” function. M-estimators are categorized into three types [8] by the behavior of  $\psi(u) = \rho'(u)$ . We study one of each type. *Monotone* M-estimators (Figure 2a), such as Huber’s [9, Chapter 7], use non-decreasing, bounded  $\psi(u)$  functions. *Hard redescenders* (Figure 2b), such as Hampel’s [5], force  $\psi(u) = 0$  for  $|u| > c$ . *Soft redescenders* (Figure 2c), such as Mirza and Boyer’s [2], force  $\psi(u) \rightarrow 0$  as  $|u| \rightarrow \infty$ . The three robust loss functions shown in Figure 2 are, in order,

$$\rho_m(u) = \begin{cases} u^2, & |u| \leq c \\ 2c(|u| - c), & c < |u| \end{cases}, \quad (2)$$

$$\rho_h(u) = \begin{cases} u^2, & |u| \leq a \\ a(2|u| - a), & a < |u| \leq b \\ a(|u| - c)^2 / (b - c) + (b + c - a), & b < |u| \leq c \\ a(b + c - a), & c < |u| \end{cases}, \quad (3)$$

and

$$\rho_s(u) = (1 + f) \log(1 + u^2/f). \quad (4)$$

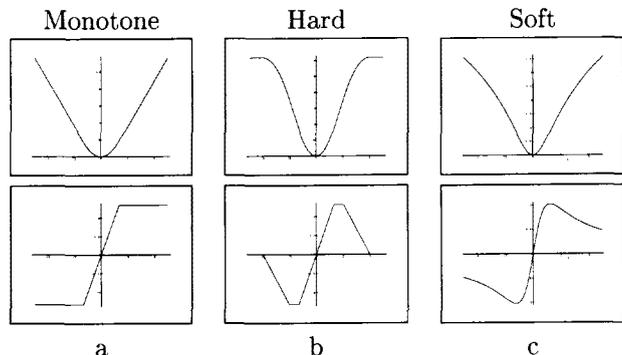


Figure 2:  $\rho(u)$ , top row, and  $\psi(u)$ , bottom row, for three M-estimators.

M-estimators typically minimize  $\sum \rho(u_i)$  using re-weighting [8] or non-linear least-squares techniques [9, Chapter 7], but only monotone  $\psi$  functions ensure a unique, global minimum.

Although M-estimators must determine  $\hat{\sigma}$ , computer vision applications often assume  $\hat{\sigma}$  is known or estimated separately. Thus, in our study, we will assume  $\hat{\sigma}$  is given and that  $\hat{\sigma} = \sigma$ . This gives the best possible robust estimator performance, since in the presence of pseudo outliers, scale estimators produce  $\hat{\sigma} > \sigma$ , leading to worse performance near discontinuities [16]. When  $\hat{\sigma}$  is fixed *a priori*, our M-estimators are no longer true M-estimators since they are not *scale equivariant* [6, page 259]. To reflect this, we will refer to them as “fixed-scale M-estimators.”

Hough transforms [10] and RANSAC methods [4] (following [14] we call these “fixed-band” techniques) are special cases of fixed-scale M-estimators. Although typically viewed as attempting to maximize the number of fit inliers within a pre-specified inlier band, they may be viewed instead as minimizing the number of outliers.<sup>1</sup> This gives a simple robust loss function

$$\rho_f(u) = \begin{cases} 0, & |u| \leq c \\ 1, & |u| > c. \end{cases} \quad (5)$$

As an alternative to the usual binary voting scheme, some recent Hough techniques use weighted voting [7, 13], making their optimization criteria a type of fixed-scale hard-redescending M-estimator (Figure 2b).

<sup>1</sup>Of course, they differ dramatically in their search techniques, and RANSAC may quit searching when a fit with “enough” inliers is found. The latter can result in extremely poor performance near discontinuities [16].

## 2.2 Least Median of Squares

Least median of squares (LMS), introduced by Rousseeuw [15], finds the fit minimizing the median of squared residuals. Its optimization criteria is, therefore,

$$\min_{\theta} \{ \text{median}[(z_i - \theta(\vec{x}_i))^2] \}. \quad (6)$$

LMS is implemented using random sampling techniques.

## 2.3 MINPRAN

MINPRAN searches for the fit minimizing the probability that a fit and a collection of inliers to the fit could be due to random outliers [18]. Assuming random outlier residuals are uniformly distributed<sup>2</sup> with a maximum value of  $Z_0$ , for any given fit  $\theta(\vec{x})$ , the probability  $k$  or more points could be within  $\theta(\vec{x}) \pm r$  if these points were all random outliers is

$$\mathcal{F}(r, k, N) = \sum_{j=k}^N \binom{N}{j} (r/Z_0)^j (1 - r/Z_0)^{N-j}. \quad (7)$$

The minimum of  $\mathcal{F}$  for any given fit must occur at one of the  $N$  ordered, absolute residuals,  $r_{\theta,i}$ , of the data relative to  $\theta(\vec{x})$  [18]. Thus, MINPRAN's optimization criteria is

$$\min_{\theta} \left[ \min_i \mathcal{F}(r_{\theta,i}, i, N) \right]. \quad (8)$$

Like LMS, MINPRAN is implemented using random sampling techniques.

## 3 Pseudo Outlier Breakdown and Bias

For a given robust estimator and pseudo outlier data model, our two new measures, pseudo outlier breakdown and pseudo outlier bias, compare the fit with the best "representative value" of the robust estimator's optimization criteria to the expected best fit to the inliers alone. A fit's representative criterion value results from applying the optimization criteria to the expected, ordered, absolute residuals relative to the fit. We describe these techniques for one dimensional images, denoting the single image coordinate by  $x$  and the range coordinate by  $z$ .

We model the data  $(x, z)$  as generated by a random outlier process or by noisy measurements from one of two curves  $\beta_1(x)$  and  $\beta_2(x)$ . The random outliers follow a uniform distribution both in the image domain,

<sup>2</sup>This has been generalized to any known outlier distribution [17].

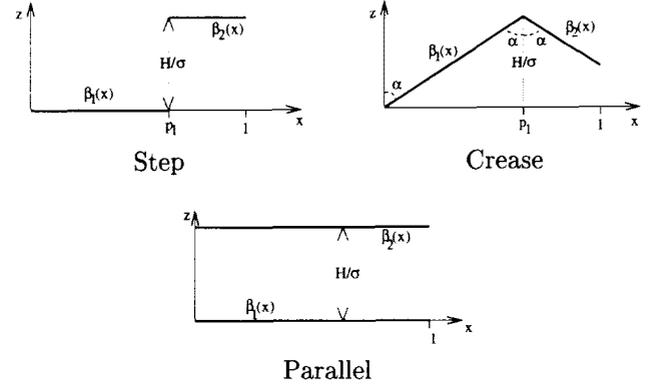


Figure 3: Parameters controlling the data models for step edges, crease edges, and parallel lines.  $H/\sigma$  is the scaled height.

which we restrict to  $[0, 1]$ ,<sup>3</sup> and in  $z$ , with a  $z$  range of  $z_0$ . Let  $[x_{i,0}, x_{i,1}]$  bound the domain of each curve and for a given  $x$ , let  $z = \beta_i(x) + \epsilon$ , where  $\epsilon$  is zero-mean, independent noise governed by gaussian density  $g(\epsilon)$ .

We use our data model to create step edges, crease edges, and parallel lines. To do this, let  $p_0$  be the probability of a random outlier; let  $(1 - p_0)p_1$  and  $(1 - p_0)(1 - p_1)$  be the probabilities points are from  $\beta_1(x)$  and  $\beta_2(x)$  respectively. For step edges and crease edges, assuming  $x$  values are governed by a uniform distribution, this implies  $x_{1,0} = 0$ ,  $x_{1,1} = x_{2,0} = p_1$ ,  $x_{2,1} = 1$  (Figure 3). For parallel lines,  $x_{1,0} = x_{2,0} = 0$  and  $x_{1,1} = x_{2,1} = 1$  (Figure 3). Requiring  $p_1 > 0.5$ , more points are from  $\beta_1(x)$  than  $\beta_2(x)$ , so points from  $\beta_1(x)$  form the inliers and points from  $\beta_2(x)$  form the pseudo outliers.

For a given data model, we can compute the expected values of the ordered, absolute residuals relative to any hypothesized fit. The pdf  $h(x, z)$  of a point  $(x, z)$  is a linear combination of the point pdfs from the random outlier process and from the data curves. The pdf for points from  $\beta_i(x)$  is

$$h_i(x, z) = \begin{cases} \frac{g(z - \beta_i(x))}{x_{i,1} - x_{i,0}}, & x_{i,0} \leq x \leq x_{i,1} \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

and the pdf for random outliers is  $h_0(x, z) = 1/z_0$ . Using  $p_0$  and  $p_1$  as above, the overall point pdf is

$$h(x, z) = p_0 h_0(x, z) + (1 - p_0) [p_1 h_1(x, z) + (1 - p_1) h_2(x, z)]. \quad (10)$$

Based on  $h(x, z)$ , the conditional cumulative distribu-

<sup>3</sup>We may limit the domain to  $[0, 1]$  because we can show [16] pseudo outlier breakdown and bias to be translation and scale invariant in both  $x$  and, assuming  $\hat{\sigma}$  is scaled appropriately,  $z$ .

tion for an absolute residual  $r$  to a fit  $\theta(x)$  is

$$F(r|\theta) = \int_0^1 \int_{\theta(x)-r}^{\theta(x)+r} h(x, z) dz dx, \quad (11)$$

and the conditional pdf of  $r$  is just  $f(r|\theta) = dF(r|\theta)/dr$ . Then, assuming  $N$  data points, we can compute the  $i$ th expected, ordered, absolute residual relative to  $\theta$  (see [3], p. 25, and [16]) as

$$E[r|\theta, i] = \int_0^\infty r N \binom{N-1}{i-1} f(r|\theta) [F(r|\theta)]^{i-1} [1-F(r|\theta)]^{N-i} dr. \quad (12)$$

For any robust optimization criteria and data model, we can compute the representative value (see the discussion in Section 1) of the optimization criteria for any fit,  $\theta(x)$ , using  $E[r|\theta, i]$  and  $F(r|\theta)$ , and then search for the fit,  $\hat{\theta}(x)$ , with the minimum representative value. For fixed-scale M-estimators the representative value is  $\sum_i \rho(E[r|\theta, i]/\hat{\sigma})$ ; for fixed-band techniques, it is  $N \cdot (1 - F(r_b|\theta))$ ; for LMS, it is the expected median absolute residual,  $E[r|\theta, [N/2]]$ ; for MINPRAN, it is  $\min_i \mathcal{F}(E[r|\theta, i], i, N)$ . We search for  $\hat{\theta}(x)$  numerically, using multiple appropriate initial  $\theta(x)$  when the optimization criteria is non-convex.

We can now define pseudo outlier breakdown and bias. Since  $\beta_1(x)$  is the expected fit to the inliers, we define *pseudo outlier bias* as the least-squares distance between  $\hat{\theta}(x)$  and  $\beta_1(x)$ , normalized by  $\sigma$ . We say that *pseudo outlier breakdown* occurs when  $\hat{\theta}(x)$  is a bridging fit. Bridging fits intersect either both curve segments (see Figure 1b) or neither. We use pseudo outlier breakdown for step edges and parallel lines since  $\hat{\theta}(x)$  is generally either  $\beta_1(x)$  or a bridging fit. We use pseudo outlier bias for crease edges, since  $\hat{\theta}(x)$  is always a bridging fit, but the distance between  $\hat{\theta}(x)$  and  $\beta_1(x)$  differs substantially depending on the robust optimization criteria.

## 4 Analysis Results

Restricting the possible fits  $\theta(x)$  to linear functions, we use pseudo outlier breakdown and bias to analyze robust optimization criteria on step edges, crease edges, and parallel lines as modeled in Section 3. Because pseudo outlier breakdown and bias are translation and scale invariant, we parameterize these discontinuities, even for crease edges, using only  $p_1$ , the fraction of points from  $\beta_1(x)$ , and  $H/\sigma$ , the scale normalized discontinuity height (Figure 3).

Figures 4 through 6 plot pseudo outlier breakdown and bias for three different types of fixed-scale M-estimators, for fixed-band techniques, for LMS, and

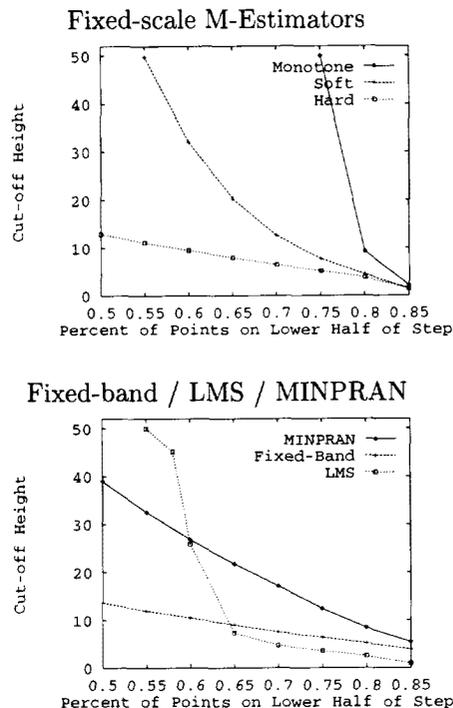


Figure 4: Pseudo outlier breakdown cutoffs for step edges. ( $H/\sigma = 50$  is the worst shown.) Below the cut-offs  $\hat{\theta}(x)$  is a bridging fit. The top plot shows cut-offs for three Fixed-scale M-estimators. The bottom plot shows cut-offs for MINPRAN, fixed-band techniques, and LMS.

for MINPRAN. The pseudo outlier breakdown plots show, as a function of  $p_1$ , the smallest  $H/\sigma$  for which  $\hat{\theta}(x)$  is a bridging fit. The pseudo outlier bias plots show, for  $p_1 = 0.6$ , the bias as a function of  $H/\sigma$ . For each of the plots, the additional data model parameters, which minimally influence the results, are  $p_0 = 0.1$ ,  $z_0 = 100\sigma$ , and  $N = 100$ . The estimator tuning parameters for fixed-scale M-estimators (equations 2 through 4) are taken from the literature ( $c = 1.345$  for  $\rho_m$  [8],  $a = 1.31$ ,  $b = 2.04$ ,  $c = 4.00$  for  $\rho_h$  [5], and  $f = 1.5$  for  $\rho_s$  [2]). For fixed-band techniques (equation 5),  $c = 2.5$ .

Overall, although the hard-re-descending fixed-scale M-estimator and LMS (when  $p_1 \geq 0.6$ ) give the best results, all estimators' pseudo outlier breakdown and bias performances are disappointing, especially since we are examining robust estimators under the best possible conditions:  $\hat{\sigma} = \sigma$  for estimators requiring a scale parameter, and  $\hat{\theta}(x)$  is (approximately) at a global minimum of the robust optimization criteria.

We can make several additional observations about

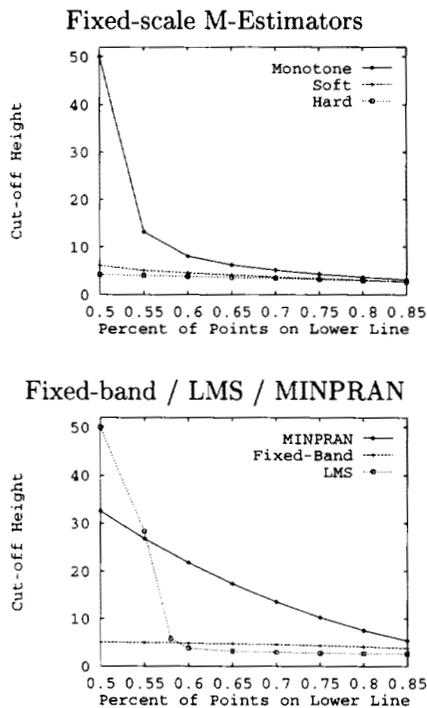


Figure 5: Pseudo outlier breakdown cutoffs for parallel lines. ( $H/\sigma = 50$  is the worst shown.) Below the cut-offs  $\hat{\theta}(x)$  is a bridging fit. The top plot shows cut-offs for three fixed-scale M-estimators. The bottom plot shows cut-offs for MINPRAN, fixed-band techniques, and LMS. Pseudo outlier breakdown cut-offs are lower than for step edges because bridging fits' residuals, on average, are larger for parallel lines.

the results, starting with fixed-scale M-estimators and fixed-band techniques. See [16] for results justifying the third and fourth items.

- The monotone and soft-redescending fixed-scale M-estimators optimization criteria have extremely poor pseudo outlier performance. Apparently, a finite outlier rejection point, characteristic of hard-redescending M-estimators, is necessary to tolerate, at least to some degree, the asymmetry and coherent structure of pseudo outliers.
- The hard-redescending fixed-scale M-estimator's optimization criterion is better than the binary, fixed-band optimization criteria because it retains the statistical efficiency of least-squares for small residuals. The difference between these is most apparent in the pseudo outlier bias measures for crease edges.

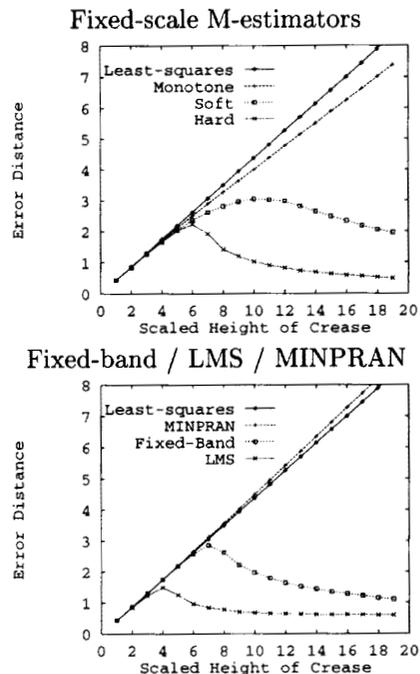


Figure 6: Pseudo outlier bias for crease edges when  $p_1 = 0.6$ . The horizontal axis is the scaled crease height, and the vertical axis is the  $L^2$  distance between  $\beta_1(x)$  and  $\hat{\theta}(x)$ . The top plot shows three fixed-scale M-estimators. The bottom plot shows MINPRAN, fixed-band techniques, and LMS. The pseudo outlier bias curve for least-squares is included in each figure. As  $p_1$  increases, each curve gradually flattens.

- Reducing the tuning parameters to narrow  $\rho_h$  (equation 3) for the hard-redescending fixed-scale M-estimator lowers its pseudo outlier breakdown cut-offs and reduces its pseudo outlier bias [16]. This justifies the lower parameter values commonly used in vision [1, 11], although pseudo outlier breakdown and bias still occur for small magnitude discontinuities.
- When  $\sigma$  must be estimated along with  $\theta(x)$ , M-estimator performance is much worse than that of fixed-scale M-estimators because  $\hat{\sigma}$  is generally over-estimated making discontinuity magnitudes appear smaller relative to  $\hat{\sigma}$  [16].
- Obtaining the optimum fit using a hard-redescending fixed-scale M-estimator is difficult because  $\rho_h$  (equation 3) is non-convex. In general, to find the global optimum, expensive, non-gradient search techniques, such as the Hough transform or random sampling, must be used.

Of the two estimators that do not require  $\hat{\sigma}$ , LMS

and MINPRAN, LMS has the better pseudo outlier performance, at least when  $\beta_1(x)$  has enough inliers, i.e. when  $(1-p_0)p_1 > 0.5$ . MINPRAN's performance, which is quite poor, can be improved dramatically by a minor change that allows it to extract the two best disjoint (no shared inliers) fits and compare these to the single best fit, selecting one fit or two depending on which appears less random [17]. This change makes MINPRAN as effective as the hard-re-descending fixed-scale M-estimator.

## 5 Conclusions

By defining pseudo outlier breakdown and bias, new robust estimator performance measures that are based on expected data values, and using them to study robust estimator optimization criteria on data containing inliers, random outliers, and pseudo outliers, we have shown that robust estimators can not reliably extract accurate polynomial surface patches in the neighborhood of discontinuities.

Choosing between the estimators to get the best possible performance depends on the information available about the data: when the scale parameter  $\sigma$  is known *a priori* and constant throughout the data, one should use a hard-re-descending fixed-scale M-estimator optimization criteria with reduced tuning parameters and either a random-sampling or weighted Hough transform search technique. When  $\sigma$  is unknown but the random outlier distribution is known, one should use the modified MINPRAN algorithm [17]. Finally, when neither  $\sigma$  nor the outlier distribution is known, one should use least median of squares (LMS).

Since robust estimator performance on pseudo outliers is insufficient, the pseudo outlier problem must be addressed explicitly to avoid the problems we have described. The simplest technique is to avoid image regions that may contain discontinuities until good initial fits are known in other regions [2]. This technique, however, assumes piecewise smooth data with isolated discontinuities, an assumption which can not, in general, be met. Thus, a general solution to the pseudo outliers problems will require simultaneously obtaining multiple fits when there are multiple surfaces in an image region. This is the main goal of our ongoing research.

## References

- [1] P. J. Besl, J. B. Birch, and L. T. Watson. Robust window operators. In *Proc. IEEE Int. Conf. on Computer Vision*, pages 591–600, 1988.
- [2] 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. on PAMI*, 16:987–1001, 1994.
- [3] H. A. David. *Order Statistics*. John Wiley and Sons, 1970.
- [4] M. A. Fischler and R. C. Bolles. Random Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *CACM*, 24:381–395, 1981.
- [5] F. R. Hampel, P. J. Rousseeuw, and E. Ronchetti. The change-of-variance curve and optimal re-descending M-estimators. *J. Amer. Stat. Assoc.*, 76:643–648, 1981.
- [6] F. R. Hampel, P. J. Rousseeuw, E. Ronchetti, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, 1986.
- [7] W. Hoff and N. Ahuja. Surfaces from stereo: Integrating feature matching, disparity estimation and contour detection. *IEEE Trans. on PAMI*, 11:121–136, 1989.
- [8] P. W. Holland and R. E. Welsch. Robust regression using iteratively reweighted least-squares. *Commun. Statist.-Theor. Meth.*, A6:813–827, 1977.
- [9] P. J. Huber. *Robust Statistics*. John Wiley & Sons, 1981.
- [10] J. Illingworth and J. Kittler. A survey of the Hough transform. *CVGIP*, 44:87–116, 1988.
- [11] P. Meer, D. Mintz, A. Rosenfeld, and D. Y. Kim. Robust regression methods for computer vision: A review. *Int. J. of Computer Vision*, 6:59–70, 1991.
- [12] D. Mintz, P. Meer, and A. Rosenfeld. Analysis of the least median of squares estimator for computer vision applications. In *Proc. IEEE Conf. on CVPR*, pages 621–623, 1992.
- [13] J. Princen, J. Illingworth, and J. Kittler. Hypothesis testing: A framework for analyzing and optimizing Hough transform performance. *IEEE Trans. on PAMI*, 16:329–341, 1994.
- [14] G. Roth and M. D. Levine. Extracting geometric primitives. *CVGIP: Image Understanding*, 58:1–22, 1993.
- [15] P. J. Rousseeuw. Least median of squares regression. *J. Amer. Stat. Assoc.*, 79:871–880, 1984.
- [16] C. V. Stewart. Expected performance of robust estimators near discontinuities. Technical Report 94-10, Department of Computer Science, Rensselaer Polytechnic Institute, 1994.
- [17] C. V. Stewart. MINPRAN: A new robust estimator for computer vision (accepted for publication). *IEEE Trans. on PAMI*, 1994.
- [18] C. V. Stewart. A new robust operator for computer vision: Theoretical analysis. In *Proc. IEEE Conf. on CVPR*, pages 1–8, 1994.