Robust High Breakdown Estimation and Consensus

Peter Meer
Department of Electrical and Computer Engineering
Rutgers University
Piscataway, NJ 08855-0909

Abstract

Application of robust regression methods to computer vision problems becomes increasingly popular. These methods make use of a priori available information through the assumed model from which, however, deviations are allowed. Thus robust methods are more powerful than nonparametric techniques which, in order to avoid erroneous assumptions, do not involve models at all. We describe the least median of squares estimator and show that for computer vision data its staightforward use may not yield the desired results. Most of the artifacts can be avoided by using a somewhat different, consensus based approach. We present the principle behind the consensus paradigm.

Keywords: Robust estimation, Least Median of Squares, Consensus

Introduction

In the last two decades, to avoid the pitfalls of least squares, robust estimation techniques were developed by statisticians. Recently these robust techniques have become popular in computer vision. See Meer *et al.* (1990b) for a survey of robust techniques as applied in computer vision. Some of the robust operators can tolerate a discontinuity (edge) and recover the model corresponding to the absolute majority of the data in the processing window. The least median of squares (LMedS) estimator, proposed by Rousseeuw in 1984, is one of this class and is the most frequently used in computer vision. However, the LMedS estimator was developed for statistical applications and its utility in solving computer vision problems cannot be taken for granted. In Section 3 we describe LMedS and show its limitations for noisy data. This limitations can be avoided by a *consensus* based technique which retains the desirable properties of both the least squares and the least median of squares estimators. In Section 4 we present the consensus principle.

The Weighted Least Squares Estimator

The weighted least squares (WLS) estimator is a frequently used tool in computer vision. We only mention here the properties used later in the paper. For proofs and/or more detailed discussions see a textbook on linear estimation. The noise corrupting the linear model used in WLS is assumed to be of zero mean. In this case the estimated parameters are unbiased, i.e., their expected value is correct. The WLS estimates become minimum-variance estimates by proper choice of the weights. However, it can be shown that for the small window operators usually employed in computer vision the achieved accuracy does not suffices (Meer *et al.*, 1990a).

The implicit assumption behind WLS is that the entire data set can be characterized by only *one* parameter vector β. For piecewise data, a case frequently met in computer vision but less common in statistics, this is not true. In piecewise data a model discontinuity is also present and (at least) two parameter vectors are required to describe the entire data set. For example, changes in reflectance or depth yield piecewise data.

It is well known that least squares estimation cannot handle piecewise data. This is caused by the violation of the zero mean noise assumption. The data not represented by the model has to be regarded as corrupted by an additional nonzero mean noise process. It is shown in the next section that model estimation for piecewise data can be achieved only by high breakdown point robust techniques which tolerate nonzero mean noise.

Least squares polynomial surface estimation of piecewise constant data (i.e., containing a step edge) has a property of importance for us. Let the step edge in the data be of height h and corrupted with noise having standard deviation σ. We estimate the surface in a window centered on the discontinuity. The simplest, degree-0 model is that of a constant plane (the number of unknown parameters $p = 1$). The least squares estimated value yields a plane lying between the two original horizontal surfaces that form the step edge. When a planar, degree-1 model ($p = 3$) is used, the estimated plane lies across the edge and intersects both surfaces.

Both fits are erroneous and do not correspond to either of the original surfaces. However, up to moderate signal-to-noise (SNR) ratios, when the standard deviation of the residuals (the noise's standard deviation estimate) is computed the degree-0 model yields an increase relative to σ while the degree-1 model may not. Thus, when a window operator performing degree-0 least squares estimation (averaging) slides along piecewise constant noisy data, the standard deviation estimate increases in the neighborhoods of step edges. The increase is significant enough to be used for coarse detection of the presence of a model discontinuity (edge). Thus a reliable indication of the presence of a discontinuity is obtained. This result is widely applied for adaptive filtering in the signal processing literature. For an application in computer vision see Park and Meer (1990).

To conclude this section we emphasize two observations:

- When all the data belongs to one model corrupted with zero mean noise the least squares estimates are unbiased but may have a large standard deviation.
- At moderate signal-to-noise ratios the degree-0 least squares estimate of the noise's standard deviation can be used as warning about the presence of a step edge in the data.

The Least Median of Squares Estimator

In this section we discuss a robust estimator which can handle weakly corrupted piecewise data. For convenience we describe the one-dimensional case; extension to multi-dimensions is immediate.

The ability of an estimator to handle severely corrupted data is captured by its *breakdown point* ε^*. This is the smallest fraction of the data which can yield arbitrary estimate values. For example, the degree-0 least squares estimate (the *mean* of n data points, z_i) has $\varepsilon^* = 1/n$, since one large valued erroneous point already compromises the result. The asymptotical breakdown point of the degree-0 LS estimator thus is 0. The zero breakdown point property is common for all least squares estimators (Hampel *et al.*, 1986, p.328).

Two types of severely corrupted data can be distinguished. *Outliers* are data points with values $z_{i,j}$ which cannot be represented by the assumed model. Thus noiseless piecewise data can always be described by one model and a set of outliers. For example, one of the planes forming a roof edge is taken as the model and the points belonging to the other are then regarded as outliers relative to that model. Once zero mean noise is also present, however, the dichotomy between the model and the outliers may not be easy to establish. (We return to this subject later in this section.)

Leverage points are data points corresponding to outlying regressor variable values $x_k(i,j)$. These points are not necessarily "bad" since their values may be close to the value predicted by the model. However, if a data point is far away from the rest and cannot be accounted for by the model, it may exert an increased influence on the estimate. The sparse data in some computer vision applications (for example, stereo) may contain leverage points which compromise the result in spite of the fact that their value is within the range of the data. Detection of leverage points requires special techniques. See for example Rousseeuw and van Zomeren (1990).

The breakdown point of an estimator cannot exceed 0.5 without making use of a priori information about the data. Indeed, at least half of the data should be represented by the model to be the unique solution of the estimation. The *median,* the degree-0 least absolute deviations estimator (belonging to the L_1 family of estimators), tolerates close to half the data being severely corrupted. At the limit, the median has $\varepsilon^* = 0.5$. It can be shown, however, that higher degree surface fitting in L_1 is sensitive to leverage points and it has $\varepsilon^* = 0$ (Hampel *et al.*, 1986, p.328).

Another degree-0 estimator is the *mode* of the data. For a continuous probability distribution function (p.d.f.) the mode is the most probable variable value, i.e., the maximum of the p.d.f. (Without loss of generality we can consider a unimodal p.d.f. for the moment.) Assume that n outcomes were obtained from the p.d.f. and they are ordered by increasing values. We first compute the half-length of a window containing $N \le n/2$ points in all the $n - N+1$ positions along the ordered sequence. (The half-length is defined as half the difference between the data point values at locations i and $i+N-1$ in the ordered sequence.) The midpoint of the shortest window is taken as the mode since the maximum of the p.d.f. implies the most outcomes in an interval.

Let the window size equal to half the data size, i.e., $N = \lfloor n/2 \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function. Then it can be shown that the mode $\hat\beta_0$ minimizes the median of the squared residuals, i.e., satisfies

$$\min_{\hat\beta_0} \operatorname*{med}_i (z_i - \hat\beta_0)^2 \qquad (1)$$

and that the value of (1) is the squared half-length of the window yielding the mode (Rousseeuw and Leroy, 1987, p. 166). The use of squares instead of absolute values assures the uniqueness of the solution for n an even number *(ibid,* p. 170). The presence of the median in (1) makes the mode a robust degree-0 estimator with breakdown point close to 0.5.

The minimization problem (1) can be generalized to an arbitrary model $\hat\beta$ generating the residuals r_i

$$\min_{\hat\beta} \operatorname*{med}_i r_i^2 . \qquad (2)$$

The expression (2) defines the *least median of squares* (LMedS) estimator introduced in statistics by Rousseeuw in 1984. The already mentioned book of Rousseeuw and Leroy (1987) gives an excellent practical analysis of the estimator. The finite sample breakdown point of the LMedS estimator is

$$\varepsilon^* = \frac{\lfloor n/2 \rfloor - p + 2}{n} \qquad (3)$$

and asymptotically yields 0.5.

Since (2) has no analytical solution, to find the LMedS estimates a numerical technique, projection pursuit, is employed.

The LMedS Algorithm

Assumptions: The model is defined by the parameters β_k, $k = 0, \ldots, (p-1)$. The data contains n points with a fraction $\varepsilon < 0.5$ of outliers.

Step 1. Randomly select a p-tuple Π from the data.

Step 2. For this p-tuple compute the parameter values $\hat\beta_k(\Pi)$ by solving the p equations. Note that if a linear model is assumed the p-tuple must yield a full rank coefficient matrix for the system.

Step 3. Retain all parameters but $\hat\beta_0(\Pi)$. Project the data into the β_0 subspace by computing

$$\alpha_i(\Pi) \triangleq z_i - \sum_{k=1}^{p-1} \hat\beta_k(\Pi) x_k(i) \qquad i = 1, \ldots, n . \qquad (4)$$

Step 4. Find the mode of the $\alpha_i(\Pi)$ sequence and allocate it to $\hat\beta_0(\Pi)$. Store the corresponding shortest half-window size

$$\delta(\Pi) = [\min_{\hat\beta(\Pi)} \operatorname*{med}_i r_i^2]^{1/2} . \qquad (5)$$

Step 5. Repeat Steps 1 to 4 q times. The final LMedS estimates are from the p-tuple yielding $\delta = \min_{\Pi} \delta(\Pi)$.

The search procedure (Step 5) has the goal of finding at least one p-tuple containing only data points represented by the model, i.e., *inliers*. Projection of such a p-tuple into the β_0 space (Step 3) can yield at most εn outliers. The mode detection procedure (Step 4) then assures the recovery of $\hat{\beta}_0(\Pi)$, and that the entire parameter vector satisfies (2).

The random sampling (Step 1) reduces the complexity of the LMedS algorithm to a manageable amount of computation. Mode computation per p-tuple requires $O(n \log n)$ operations and there are $O(n^p)$ possible p-tuples. Let the tolerated probability of error in random sampling be $Q \ll 1$. Then the probability that for q independently selected p-tuples at least one contains only inliers is

$$1 - [\,1 - (1-\varepsilon)^p\,]^q \le 1 - Q\ . \qquad (6)$$

From (6) q can easily be computed. For example, when a degree-1 model ($p=3$) is fit to piecewise data with $\varepsilon=0.45$, and the tolerated probability of error is $Q=0.01$, choosing $q=26$ different 3-tuples suffices to find the three LMedS estimates. Note that q does not depend on the data size. Since the LMedS estimates correspond to one of the p-tuples, the only characteristic of interest about the structure of the data is ε.

A "hidden" assumption behind the LMedS estimator is that the inliers are only weakly corrupted by i.i.d. Gaussian noise. That is, the parameter values obtained from a p-tuple containing only inliers are close to the correct ones. The LMedS algorithm also returns a robust estimate of the noise's standard deviation:

$$\hat{\sigma} = 1.4826 \left[1 + \frac{5}{n-p} \right] \delta \qquad (7)$$

where the term $5/(n-p)$ is the finite sample size correction (*ibid*, p. 202) and 1.4826 is the correction factor for median based Gaussian noise standard deviation estimates. Once the noise corrupting the model is defined (zero mean normal with variance $\hat{\sigma}^2$) the data can be classified into inliers and outliers. The weights

$$w_i = 1 \quad \text{if} \quad |r_i| \le 2.5\hat{\sigma} \qquad (8)$$

mark the inliers with about a 0.95 confidence level (the uncertainty is due to the finite data size). If the nature of the noise is known a priori other correction factors and thresholds can be computed. The LMedS algorithm and the inlier/outlier dichotomy procedure, however, become unreliable when significant noise corrupts the data. This problem is discussed below.

An optimization principle similar to the one used in the LMedS algorithm was already proposed for computer vision applications by Fischler and Bolles (1981) as the Random Sample Consensus (RANSAC) paradigm. In RANSAC the model derived from a p-tuple determines the set of data points agreeing with it within a given tolerance limit. If the number of points in the set exceeds a threshold, no more p-tuples are drawn. Thus RANSAC requires the a priori information of a tolerance limit and a cardinality threshold. LMedS automatically selects the model representing at least half the data by minimizing a quality measure of the fit, the median of the squared residuals. For a more detailed comparison of the two techniques see Meer *et al.* (1990b).

The high breakdown point of the LMedS estimator makes it an ideal operator for handling piecewise data. The estimator identifies the model corresponding to the majority of the points in the set and discriminates the points not belonging to it, i.e., the outliers. The LMedS algorithm (as presented above) was applied to several computer vision problems. Kim *et al.* (1989), Roth and Levine (1990), Sinha and Schunck (1990) have used it in surface reconstruction; Kumar and Hanson (1989) in camera position estimation; Tirumalai *et al.* (1990) in dynamic stereo.

We have already emphasized the necessary condition for obtaining "good" LMedS estimates, the existence of at least one p-tuple from which reliable model parameters can be extracted. When significant zero mean noise corrupts *all* the data points such a p-tuple may not exist. In this case a p-tuple containing only noisy inliers yields, unbiased estimates (since we are using a least square type technique for the computation) but with large variance. The noise can also erase the clear dichotomy between data points belonging to the model and outliers.

Consider again piecewise constant data containing a step edge of amplitude h (ε close to 0.5) corrupted with zero mean noise taking values in the range $(-a, a)$. This definition is for convenience only; for Gaussian noise a could be, say, 3σ. Suppose also for the moment that $a \le h/2$.

First we use the degree-0 LMedS estimator, i.e., compute the mode of the data. Assume that we have succeeded in recovering the uncorrupted value of the horizontal surface corresponding to the majority of the pixels. Then the squared residuals are distributed as follows: slightly more then half in the interval $(0, a^2)$ and the rest in the interval $[(h-a)^2, (h+a)^2]$. The estimate of $\hat{\sigma}$, the model noise standard deviation, is proportional to the square root of the median of squared residuals, and thus it will exceed a. But a is the range of the noise and therefore $\hat{\sigma}$ is a strong overestimate. Large $\hat{\sigma}$ reduces the weights and outliers may be classified as inliers.

When a degree-1 LMedS estimator is applied to the noisy piecewise constant data the large variance of the 3-tuple based estimates makes any plane orientation possible. Assume that we have chosen two 3-tuples, one generating the correct horizontal fit, the other yielding a plane tilted at about 45 degrees across the edge. The latter fit is similar to what a least squares estimator would have recovered from the window. If significant noise is present (a is close to $h/2$), most of the squared residuals for the tilted plane fit are between 0 and a^2. The median of this residual sequence is significantly lower than the median of the sequence obtained when the horizontal fit is employed (see above). The LMedS criterion (2) seeks the minimum and therefore the tilted plane is preferred. Note that tilted planes are always obtained from some 3-tuples since any sample containing pixels from both surfaces can yield one. The estimator will choose the correct, horizontal plane only when the amplitude of the corrupting noise is very small relative to the amplitude of the step, that is, the SNR is very high. The correct fit, however, implies an overestimate of σ as was shown above.

For noisy data with a discontinuity the mode detection procedure can also fail. Such data yields a bimodal distribution. The significant noise level spreads the values across the entire range of the data and no clear bimodality can be seen. The mode detection procedure finds that the shortest window corresponds to a value between the two surfaces forming the original step edge.

The severity of the artifacts of the LMedS estimator for piecewise constant data were investigated by a simulation study (Mintz *et al.*, 1990). We have shown that application of LMedS to noisy piecewise constant data with a large fraction of outliers can result in failure.

- The high breakdown point property of the LMedS estimator no longer is satisfied and performance similar to least squares is often obtained.
- The LMedS estimator cannot recover the correct model when this is misspecified, i.e., non-essential parameters are introduced.
- The inlier/outlier dichotomy can become erroneous because of overestimation of the noise's standard deviation.

In the next section we propose a new approach toward computer vision problems based on seeking consensus among local processes.

The Consensus Paradigm

The principle behind the LMedS estimator (and the RANSAC paradigm) can be described as follows:

Compute a candidate model based on a randomly chosen small subset of the data. Apply this model to all the data. Compute a global quality measure for the model. Optimize the quality measure by repeating the procedure several times.

To have reliable LMedS estimates it is necessary that at least one candidate model carries the correct parameter values. We have shown that when this condition is not satisfied the LMedS estimator loses most of its attractive properties. The high breakdown point of an estimator can be interpreted as seeking the consensus in the data about the assumed model. Optimality in LMedS is achieved by minimizing the criterion (2) and in RANSAC by maximizing the number of inliers. We can regard this optimality also as a *consensus* about the model representing the majority of data points.

However, consensus can be achieved in other ways. Assume that the majority of the data obey a model of known structure but unknown parameters. The agreement with the model is characterized by the quantity γ.

1. Define a simple local process which provides unbiased estimates of γ whenever the data agrees with the assumed model.
2. Build a global distribution of γ.
3. Find by a robust procedure the expected value of γ for agreement with the model, Γ.
4. Use Γ to select the reliable local processes.
5. Estimate the model parameters with these processes.

The local estimates are obtained by a least squares based procedure which has the advantage of being computationally inexpensive and providing unbiased estimates. The large variance of the estimates is no longer of concern since we use their global distribution. The robust procedure analysing this global distribution is based on the mode seeking algorithm described in the previous section. The mode is a high breakdown point estimator and thus close to half the local estimates can be erroneous while the value of Γ indicating agreement with the model is still correctly recovered. The spread of the γ estimates around Γ is charaterized by the robust standard deviation (7). The "inliers" of this distribution thus mark the local processes where agreement with the model was obtained. The applicability of the described paradigm is very large. We describe succintly two applications.

In *multiresolution adaptive least squares* image smoothing (Park and Meer, 1990) local fits for a constant plane are computed at every pixel. Several neighborhood sizes $7\times7, 5\times5, 3\times3$ are used. For given neighborhood size the standard deviation of the residuals is used as γ. Then $\gamma\leq\Gamma$ indicates the pixels whose smoothed value can be taken from the given neighborhood since the neighborhood falls on a homogeneous region not containing an edge. The analysis proceeds from large to small neighborhoods.

In the *consensus by decomposition* (CBD) estimators (Mintz *et al.* 1990) a candidate model is computed for a small neighborhood of every data point. Each model parameter is mapped into a separate parameter-image based on the same sampling structure as the input. Note that, these parameter-images are that of a noisy step-edge. The consensus technique is applied for each parameter-image separately. The standard deviation of the residuals from a constant are used as γ. The points around a discontinuity (if exist) are eliminated by mapping the outliers of the γ distribution into the parameter-image. A second mode seeking procedure recovers the robust parameter estimate. The CBD estimation proceeds from the highest degree coefficients of a polynomial surface model toward the intercept. The necessary condition for a reliable CBD estimate is that the majority of the candidate models carry unbiased parameter estimates. This condition is much weaker than the one required by LMedS and makes possible the elimination of the latter's artifacts.

Conclusion

We have described a new principle to approaching computer vision problems. Whenever no clear dichotomy between "good" and "bad" data exist, the robust high breakdown methods from statistic may fail. The decomposition of the problem into non-robust local estimation procedures yields unbiased distributions for the sought estimates. When these distributions are analysed with a simple mode seeking algorithm all the desired properties of high breakdown estimation are preserved.

References

[1] M.A. Fischler and R.C. Bolles (1981): Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* **24**, 381–395.

[2] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel (1986): *Robust Statistics: An Approach Based on Influence Functions*. John Wiley & Sons: N.Y.

[3] D.Y. Kim, J.J. Kim, P. Meer, D. Mintz and A. Rosenfeld (1989): Robust computer vision: A least median of squares based approach. In *Proceedings of the DARPA Image Understanding Workshop,* Palo Alto, CA, May 1989, 1117–1134.

[4] R. Kumar and A.R. Hanson (1989): Robust estimation of camera location and orientation from noisy data having outliers. In *Proceedings of the IEEE Workshop on Interpretation of 3D Scenes,* Austin, TX, November 1989, 52–60.

[5] P. Meer, D. Mintz and A. Rosenfeld (1990a): Least median of squares based robust analysis of image structure. In *Proceedings of the DARPA Image Understanding Workshop,* Pittsburgh, PA, September 1990, 231–254.

[6] P. Meer, D. Mintz, D.Y. Kim and A. Rosenfeld (1990b): Robust regression methods in computer vision: A review. To appear, *International Journal of Computer Vision.*

[7] D. Mintz, P. Meer and A. Rosenfeld (1990): Consensus by decomposition: A paradigm for fast high breakdown point robust estimation. Computer Vision Laboratory. University of Maryland, College Park, CAR-TR-525, December 1990.

[8] R.-H. Park and P. Meer (1990): Multiresolution adaptive least squares smoothing of images. Computer Vision Laboratory, University of Maryland, College Park, CAR-TR-513.

[9] G. Roth and M.D. Levine (1990): Segmentation of geometric signals using robust filtering. In *Proceedings of the 10th International Conference on Pattern Recognition,* Atlantic City, NJ, June 1990, 826–831.

[10] P.J. Rousseeuw and A.M. Leroy (1987): *Robust Regression & Outlier Detection.* John Wiley & Sons: N.Y.

[11] P.J. Rousseeuw and B.C. van Zomeren (1990): Unmasking multivariate outliers and leverage points. With comments by R.D. Cook and D.H. Hawkins; D. Ruppert and D.G. Simpson; P.J. Kempthorne and M.B. Mendel; and rejoinder. *Journal of the American Statistical Association* **85,** 633–651.

[12] S.S. Sinha and B.G. Schunck (1990): A robust method for surface reconstruction. In *Proceedings of International Workshop on Robust Computer Vision,* Seattle, WA, October 1990, 183–199.

[13] A.P. Tirumalai, B.G. Schunck and R.C. Jain (1990): Robust dynamic stereo for incremental disparity map refinement. In *Proceedings of International Workshop on Robust Computer Vision,* Seattle, WA, October 1990, 412–434.