On the use of high breakdown point estimators in the image analysis

CHRISTINE H. MÜLLER

April 1998

Abstract. For eliminating noise from an image, Meer et al. (1990, 1991) have proposed to use high breakdown point estimators and in particular to use the least median of squares (LMS) estimator for linear regression. But, the highest breakdown points are attained by other estimators as the Cauchy estimator and some least trimmed squares (LTS) estimators. Therefore, in this paper the behavior of the Cauchy estimator and the LTS estimator in image analysis is studied and compared with that of the LMS estimator and the least squares (LS) estimator. For that, test images with 0%, 14%, 30%, 44% and 49% noise are used. It turns out that the LTS estimator can eliminate a high amount of noise and preserves discontinuities while the Cauchy estimator behaves similar to the nonrobust LS estimator.

AMS Mathematical Subject Classification 1991: 62F35, 62J05.

Key Words: breakdown point, least median of squares estimator, least trimmed squares estimator, Cauchy estimator, noisy images.

1. Introduction

An image is given by $T \times U$ pixel values $y(t,u) \in [0,M]$, $t = 1,\dots,T$, $u = 1,\dots,U$, where often $M = 255$ is the maximum pixel value. We assume that with some probabilities the pixel values contain some error (noise) so that we have

$$y(t,u) = y^*(t,u) + \epsilon(t,u), \qquad t = 1,\dots,T,\ u = 1,\dots,U,$$

where $y^*(t,u)$ is the original signal. To eliminate the noise the pixel values $y(t,u)$ can be filtered by averaging over a $(2v+1)\times(2v+1)$ window $W(t,u)$ centered at (t,u), i.e.

$$W(t,u) := \{(t+i, u+j);\ i,j = -v,\dots,v\}.$$

For example, if the pixels are filtered (smoothed) by a mean smoother, then the filtered values $\widehat{y}(t,u)$ are given by

$$\widehat{y}(t,u) = \sum_{i=-v}^{v} \sum_{j=-v}^{v} w_{ij}\, y(t+i, u+j)$$

for $t = 3,\dots,T-2$, $u = 3,\dots,U-2$, where w_{ij} are some weights. A special mean smoother is the following

$$\widehat{y}(t,u) = x(0,0)\,(X^\top X)^{-1}\,X^\top y_{t,u},$$

where $x : \{-v,\dots,v\} \times \{-v,\dots,v\} \to I\!R^r$ is some regression function, $X = (x(-v,-v),\dots,x(-v,v),\dots,x(v,v))^\top$ and $y_{t,u} = (y(t-v,u-v),\dots,y(t-v,u+v),\dots,y(t+v,u+v))^\top$. I.e. $\widehat{y}(t,u)$ is the Gauss-Markov estimator for estimating the surface at the center (t,u) of the window $W(t,u)$ if the surface is given by the function $x(\tilde{t},\tilde{u})^\top \beta_{t,u}$ for $(\tilde{t},\tilde{u}) \in W(t,u)$. To assume such a linear model within a window makes in particular sense, since images are usually representations of real surfaces and surfaces are piecewise polynomial, even linear. Thus, for most of the windows $W(t,u)$ we can assume that the original signal has the form

$$y^*(t+i,u+j) = x(i,j)^\top \beta_{t,u} \text{ for } i,j = -v,\dots,v.$$

If some deviations from the linearity assumptions appear within a window, then these deviations can be put in the error term.

Hence, we will work with filters of the following form:

$$\widehat{y}(t,u) = x(0,0)^\top\,\widehat{\beta}_{t,u},$$

where $\widehat{\beta}_{t,u}$ is some estimator for $\beta_{t,u}$ in the window $W(t,u)$. Traditionally the least squares (LS) estimator is used for $\widehat{\beta}_{t,u}$. But, the LS estimator is known to be very sensitive against outliers and model deviations. Therefore, Meer et al. (1990, 1991) proposed to use for $\widehat{\beta}_{t,u}$ a high breakdown point estimator and in particular they proposed the least median of squares (LMS) estimator. Meer et al. assumed that the LMS estimator has a breakdown point of approximately 50%. But this holds only if the regressors

$x(-v,-v),\ldots,x(-v,v),\ldots,x(v,v)$ are in general position (see Rousseeuw and Leroy 1987). In Section 2 it is shown that the regressors are not in general position and that the highest breakdown point is attained by some other estimators as the Cauchy estimator and some least trimmed squares (LTS) estimators. In Section 3 the LTS, the Cauchy, the LMS and the LS estimator are applied to some test images with and without noise.

2. Breakdown points of LTS and Cauchy estimators

Let be $y = X\beta + z$, where $y = (y_1,\ldots,y_N)^\top \in \mathbb{R}^N$, $X = (x_1,\ldots,x_N)^\top \in \mathbb{R}^{N\times r}$, $\beta \in \mathbb{R}^r$, and $z = (z_1,\ldots,z_N)^\top \in \mathbb{R}^N$. If $\widehat{\varphi} : \mathbb{R}^N \times \mathbb{R}^{N\times r} \to \mathbb{R}^s$ is an estimator for $\varphi(\beta) = L\beta$, $L \in \mathbb{R}^{s\times r}$, then the breakdown point $\epsilon^*(\widehat{\varphi},y,X)$ of $\widehat{\varphi}$ at (y,X) is defined by (see He et al. 1990)

$$\epsilon^*(\widehat{\varphi},y,X) := \frac{1}{N}\min\left\{M \in \mathbb{N}\cup\{0\};\ B_M(\widehat{\varphi},y,X) = \infty\right\},$$

where

$$B_M(\widehat{\varphi},y,X) := \max\left\{|\widehat{\varphi}(y,X) - \widehat{\varphi}(\overline{y},X)|;\ \overline{y} \in \mathcal{Y}_M(y)\right\}$$

and

$$\mathcal{Y}_M(y) := \left\{\overline{y} = (\overline{y}_1,\ldots,\overline{y}_N)^\top \in \mathbb{R}^N;\ \mathrm{card}\{n;\ y_n \neq \overline{y}_n\} \le M\right\}.$$

To avoid the dependence on the observation vector y we also define

$$\epsilon^*(\widehat{\varphi},X) := \min\{\epsilon^*(\widehat{\varphi},y,X);\ y \in \mathbb{R}^N\}.$$

In opposite to other definitions of the breakdown point (see Donoho and Huber 1983, Rousseeuw and Leroy 1987), in the definition of He et al. (1990) the regressors x_n are unchanged. Only the observations y_n can be exchanged by outliers in this definition. Therefore this definition is appropriate for situations where the regressors are fixed. In particular in image analysis this is the case because the regressors are given by the pixel positions.

In Müller (1997), p. 58, the following upper bound for the breakdown point of a regression equivariant estimator $\widehat{\varphi}$ was shown:

$$\epsilon^*(\widehat{\varphi},y,X) \le \frac{1}{N}\left\lfloor\frac{N - \mathcal{N}_\varphi(X) + 1}{2}\right\rfloor,$$

where $\mathcal{N}_\varphi(X)$ is the maximum cardinality of a subset V of $\{x_1,\ldots,x_N\}$ so that $\varphi(\beta)$ is not identifiable at V. In particular, we have

$$\mathcal{N}_\beta(X) = \max\left\{\sum_{n=1}^N 1_V(x_n);\ V \text{ subspace of } \mathbb{R}^r\right\}.$$

Estimators which are attaining this upper bound are, for example, estimators basing on the Cauchy estimator and special least trimmed squares (LTS) estimators. A Cauchy estimator for β is defined by

$$\widehat{\beta}_C(y,X) \in \arg\min_{\beta} \sum_{n=1}^{N} \rho(y_n - x_n^\top \beta),$$

where $\rho(z) = log(1+z^2)$. A LTS estimator for β is defined by

$$\widehat{\beta}_h(y,X) \in \arg\min_{\beta} \sum_{n=l}^{h} r_{(n)}(y,X,\beta)^2,$$

where $r_n(y,X,\beta) := |y_n - x_n^\top \beta|$ and $r_{(1)}(y,X,\beta) \leq r_{(2)}(y,X,\beta) \leq \ldots \leq r_{(N)}(y,X,\beta)$. If $h = \lfloor \frac{N+2}{2} \rfloor$ and $l = \lfloor \frac{N+1}{2} \rfloor$, then the LTS estimator is the least median of squares (LMS) estimator, i.e.

$$\widehat{\beta}_h(y,X) \in \arg\min_{\beta} \operatorname{med}_{n=1,\ldots,N}(y_n - x_n^\top \beta)^2.$$

Estimators attaining the upper bound for estimating $\varphi(\beta) = L\beta$ are of the form $\varphi(\widehat{\beta})$, where $\widehat{\beta}$ is a Cauchy estimator or a special LTS estimator. This shows the following theorem.

Theorem 1. *If $\widehat{\beta}$ is the Cauchy estimator or a LTS estimator with $\left\lfloor \frac{N+\mathcal{N}_\varphi(X)+1}{2} \right\rfloor \leq h \leq \left\lfloor \frac{N+\mathcal{N}_\varphi(X)+2}{2} \right\rfloor$ and if $\mathcal{N}_\varphi(X) = \mathcal{N}_\beta(X)$, then*

$$\epsilon^*(\varphi(\widehat{\beta}), y, X) = \frac{1}{N} \left\lfloor \frac{N - \mathcal{N}_\varphi(X) + 1}{2} \right\rfloor.$$

Proof. For linear aspects $\varphi(\beta) = L\beta$ we always have

$$\epsilon^*(\widehat{\beta}, y, X) \leq \epsilon^*(\varphi(\widehat{\beta}), y, X) \leq \frac{1}{N} \left\lfloor \frac{N - \mathcal{N}_\varphi(X) + 1}{2} \right\rfloor.$$

On the other hand it was shown in Müller (1995, 1997, p. 66) for the LTS estimator and in Mizera and Müller (1996) for the Cauchy estimator that we have

$$\epsilon^*(\widehat{\beta}, y, X) = \frac{1}{N} \left\lfloor \frac{N - \mathcal{N}_\beta(X) + 1}{2} \right\rfloor \tag{2.1}$$

so that the assumption $\mathcal{N}_\varphi(X) = \mathcal{N}_\beta(X)$ provides the assertion. $\square$

There are other trimmed estimators and other M-estimators which satisfy (2.1) (see Müller 1995, 1997, Section 4.3, and Mizera and Müller 1996, respectively). But the LTS and the Cauchy estimators are the most popular estimators.

A crucial assumption in Theorem 1 is that $\mathcal{N}_\varphi(X) = \mathcal{N}_\beta(X)$ is satisfied. At least for two-dimensional linear regression, i.e. for $x(t,u) = (1,t,u)^\top$, we can show for the setup in image analysis that $\mathcal{N}_\varphi(X) = \mathcal{N}_\beta(X)$ holds.

Proposition 1. *If* $x(t,u) = (1,t,u)$, $X = (x(-v,-v), \dots, x(-v,v), \dots, x(v,v))^\top$, *and* $\varphi(\beta) = x(0,0)^\top \beta$, *then* $\mathcal{N}_\varphi(X) = \mathcal{N}_\beta(X) = 2v+1$.

Proof. Without loss of generality we can assume $v \geq 1$. The aspect $\varphi(\beta)$ is not identifiable at $\{x(-1,-v), x(-1,-v+1), \dots, x(-1,v-1), x(-1,v)\}$ so that $\mathcal{N}_\beta(X) \geq \mathcal{N}_\varphi(X) \geq 2v+1$. If we have $\{x(t_1,u_1), \dots, x(t_{2v+2}, u_{2v+2})\} \subset \{x(-v,-v), \dots, x(-v,v), \dots, x(v,v)\}$, then there exists i,j,k with $t_i = t_j$, $u_i \neq u_j$ and $t_j \neq t_k$, $u_j = u_k$, or i,j,k,l with $t_i = t_j$, $u_i \neq u_j$ and $t_k \neq t_l$, $u_k = u_l$. In the first case we have

$$\det\left((x(t_i,u_i), x(t_j,u_j), x(t_k,u_k)\right) = (t_i - t_k)(u_j - u_i) \neq 0.$$

In the second case set $a_1 := x(t_i,u_i) - x(t_j,u_j)$, $a_2 := x(t_k,u_k) - x(t_l,u_l)$, and $a_3 := x(t_i,u_i) + x(t_j,u_j)$. Then we obtain

$$\det\left((a_1, a_2, a_3)\right) = 2(t_k - t_l)(u_i - u_j) \neq 0.$$

Hence, β is identifiable in both cases, so that $2v+2 > \mathcal{N}_\beta(X) \geq \mathcal{N}_\varphi(X)$. $\square$

Regressors $x_1, \dots, x_N$ are in general position if and only if $\mathcal{N}_\beta(X) = r-1$. Hence, for $x(t,u) = (1,t,u)^\top$ and $v \geq 1$, the proposition provides that the regressors $x(-v,-v), \dots, x(-v,v), \dots, x(v,v)$ are not in general position since $\mathcal{N}_\beta(X) = 2v+1 > 2 = r-1$. Moreover, the LMS estimator does not provide the highest breakdown point. If $\widehat{\beta}_{LMS}$ is the LMS estimator and $\varphi(\beta) = x(0,0)^\top\beta$, then the breakdown point of $\varphi(\widehat{\beta}_{LMS})$ is (see Müller 1997, p. 65, 66)

$$\begin{aligned}
\epsilon^*(\varphi(\widehat{\beta}_{LMS}), X) &:= \frac{1}{N}\left(\left\lfloor \frac{N}{2} \right\rfloor - \mathcal{N}_\varphi(X) + 1\right) \\
&= \frac{2v^2}{(2v+1)^2} = \begin{cases} 0.2\overline{2} & \text{for } v = 1, \\ 0.32 & \text{for } v = 2. \end{cases}
\end{aligned}$$

If $\widehat{\beta}$ is the Cauchy estimator or the LTS estimator with $\left\lfloor \frac{N+\mathcal{N}_\varphi(X)+1}{2} \right\rfloor \leq h \leq \left\lfloor \frac{N+\mathcal{N}_\varphi(X)+2}{2} \right\rfloor$, then the breakdown point of $\varphi(\widehat{\beta})$ is according to Theorem 1 and Proposition 1

$$\begin{aligned}
\epsilon^*(\varphi(\widehat{\beta}), X) &:= \frac{1}{N}\left\lfloor \frac{N - \mathcal{N}_\varphi(X) + 1}{2} \right\rfloor \\
&= \frac{v}{2v+1} = \begin{cases} 0.3\overline{3} & \text{for } v = 1, \\ 0.4 & \text{for } v = 2. \end{cases}
\end{aligned}$$

Hence, the breakdown point of the LMS estimator is not approximately 50% and significantly smaller than that of the Cauchy and the optimal LTS estimator.

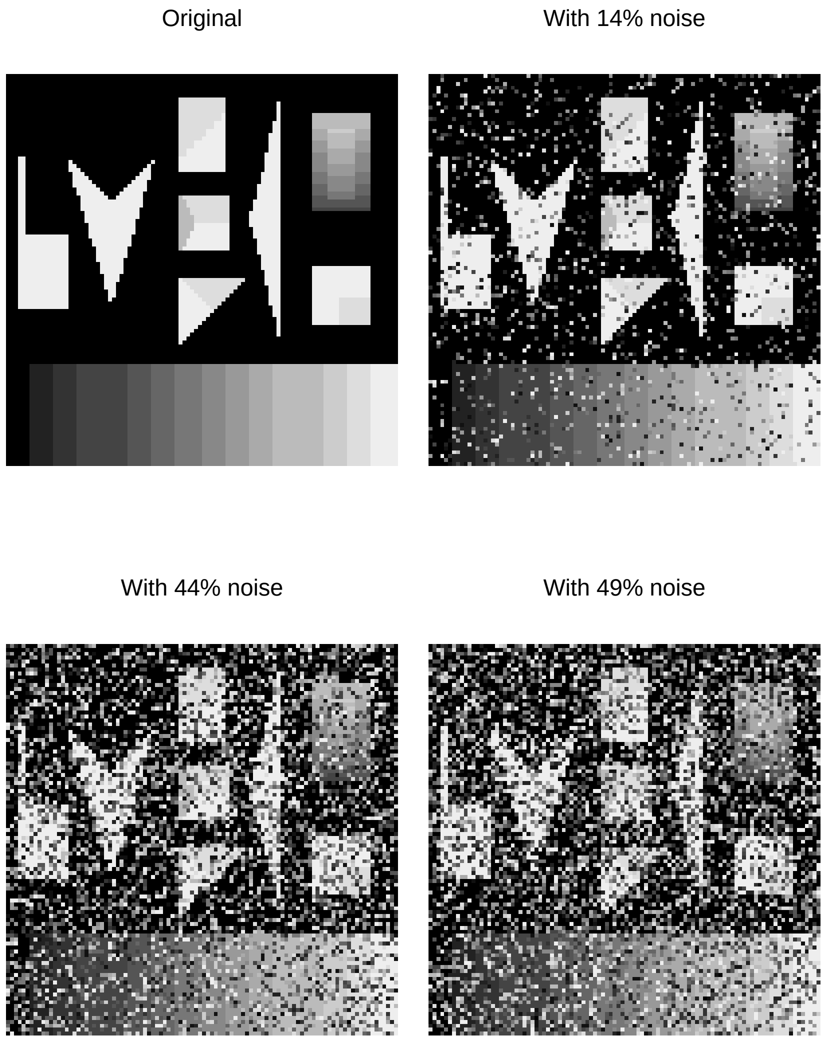

Figure 1. Original with and without noise

3. Applications on test images

As a test image an image given by Smith and Brady (1995) (downloaded from http://www.robots.ox.ac.uk/~steve/susan/susan.ps.gz) with 100×100 pixels and pixel values ranging from 0 to 255 was used. Additionally, a picture of a CD-Cover with Lee

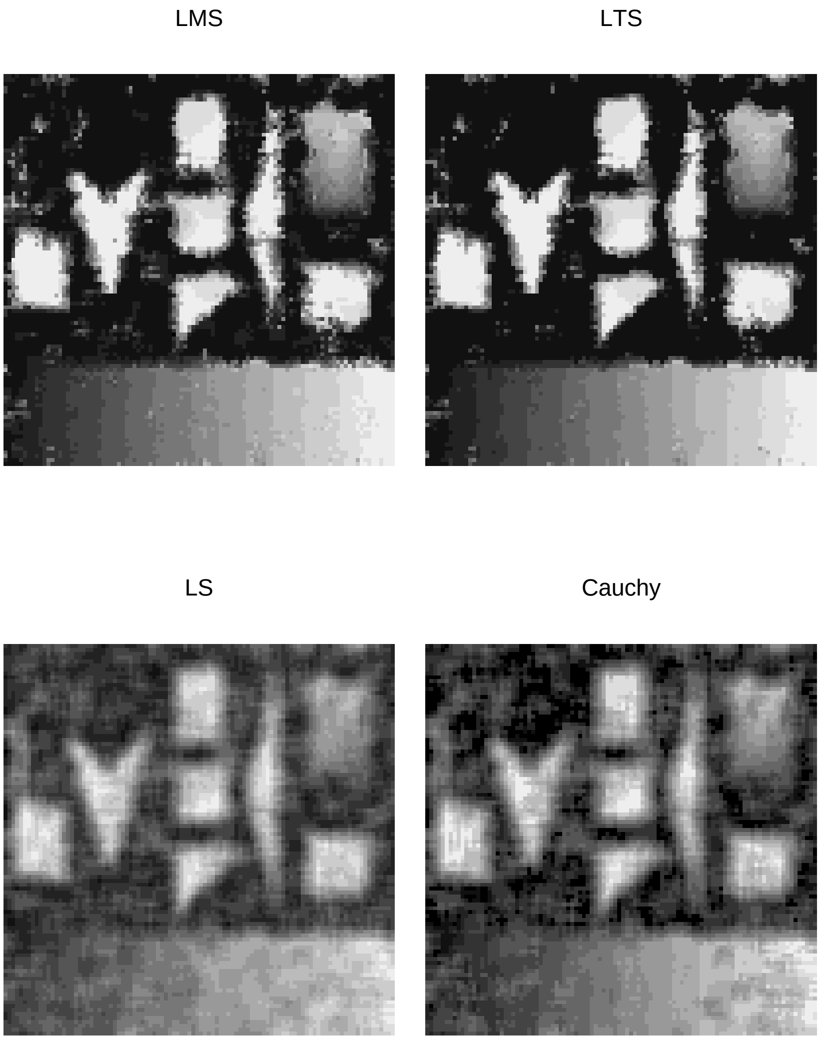

Figure 2. Estimators applied on the image with 44% noise

Perry (downloaded from http://www.rounder.com/rounder/catalog/bylabel/hbea/76/76.html) with 300×300 pixels was studied. The images were disturbed by $p = 0\%$, 14%, 30%, 44% and 49% salt and pepper noise, where in each pixel with probability p an error occurs. If an error occurs, then a value was choosen from $[0, 255]$ with uniform distribution. Although the LMS estimator is a special LTS estimator the abbreviation LTS denotes only the LTS estimator with the highest breakdown point, i.e. the LTS estimator with $h = \left\lfloor \frac{N+\mathcal{N}_\varphi(X)+1}{2} \right\rfloor$. All calculations were done with S-PLUS 3.2, where almost always 5×5 windows were used.

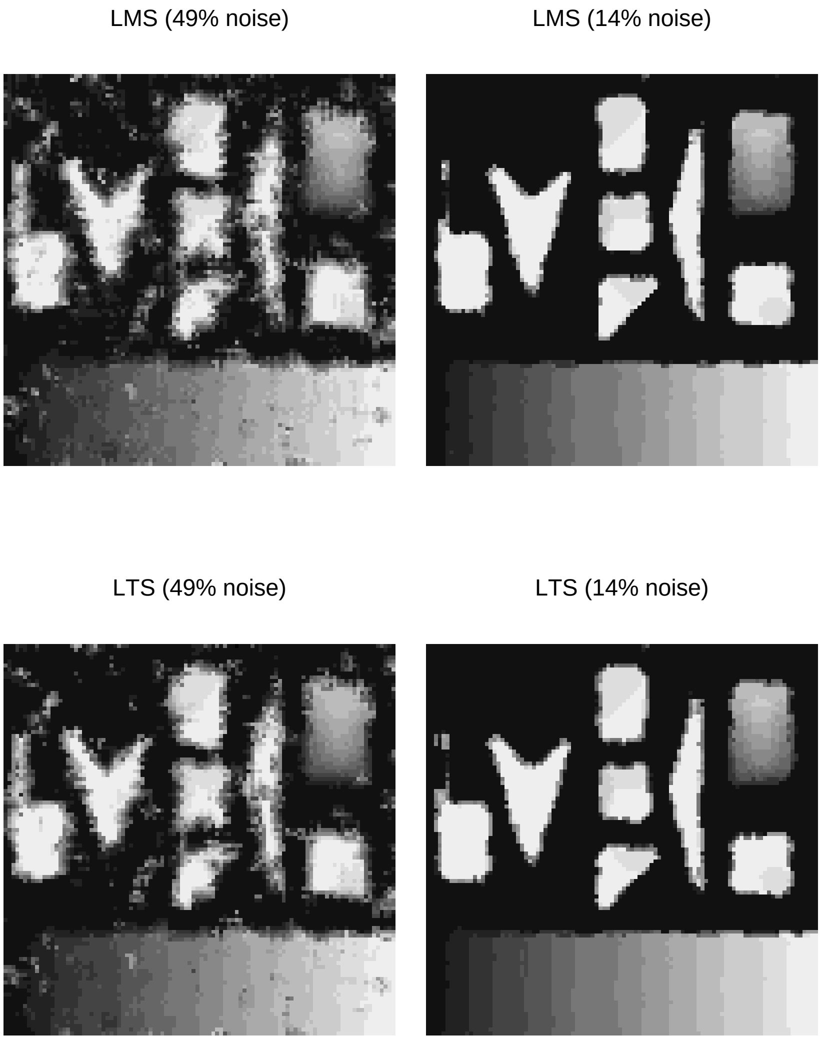

FIGURE 3. LMS and LTS applied on the image with 49% and 14% noise

4. Conclusions

Figure 2, 3 and 5 show that the LMS and the LTS estimator are able to eliminate a high amount of noise. Due to the higher breakdown point of the LTS estimator, the LTS estimator can eliminate more noise than the LMS estimator, but the differences are not

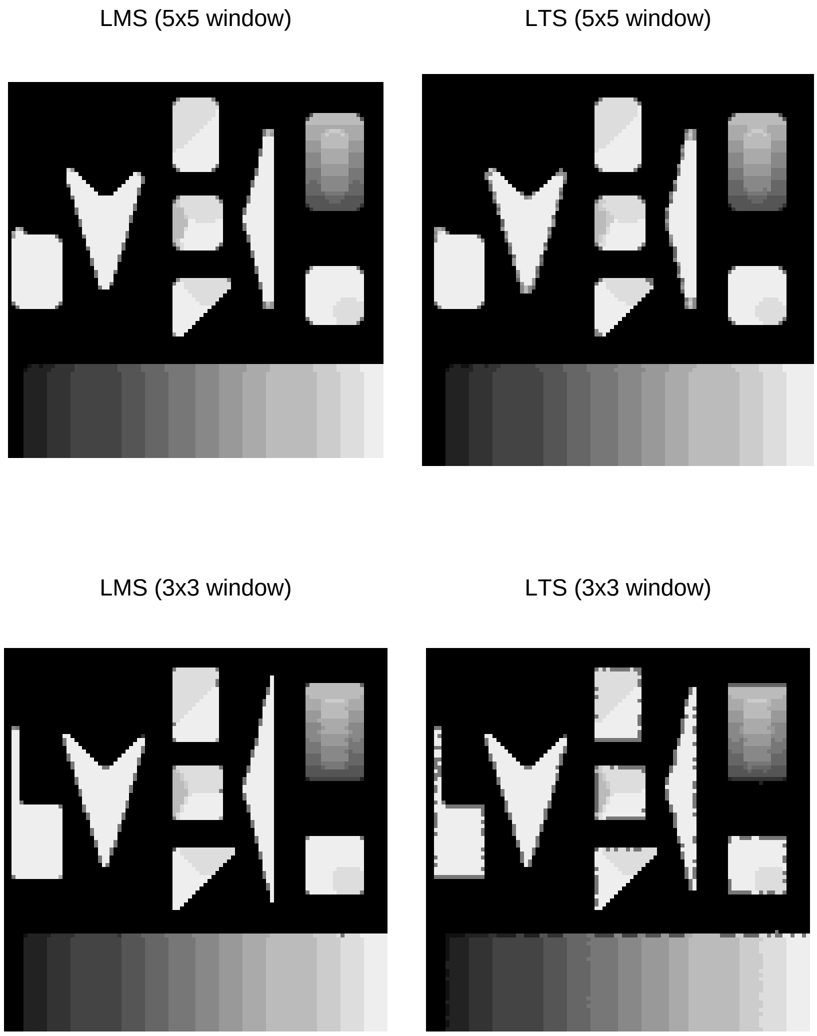

FIGURE 4. LTS and LMS applied on the original

large. In the presence of low amount of noise, say 0% or 14%, their behavior is also good (see Figure 3 and 4). In particular, they are preserving discontinuities. While with 5×5 windows the edges are preserved completely, this is not the case for the LTS estimator used at 3×3 windows (see Figure 4). The reason is that in this case the LTS estimator is not unique. Although the LMS estimator is not unique in this case as well it preserves the edges. Probably this different behavior is caused by the different programs used by S-PLUS. The program used for the LTS estimator is based on a genetic algorithm while the program for the LTS estimator is completely deterministic. In opposite to edges, the

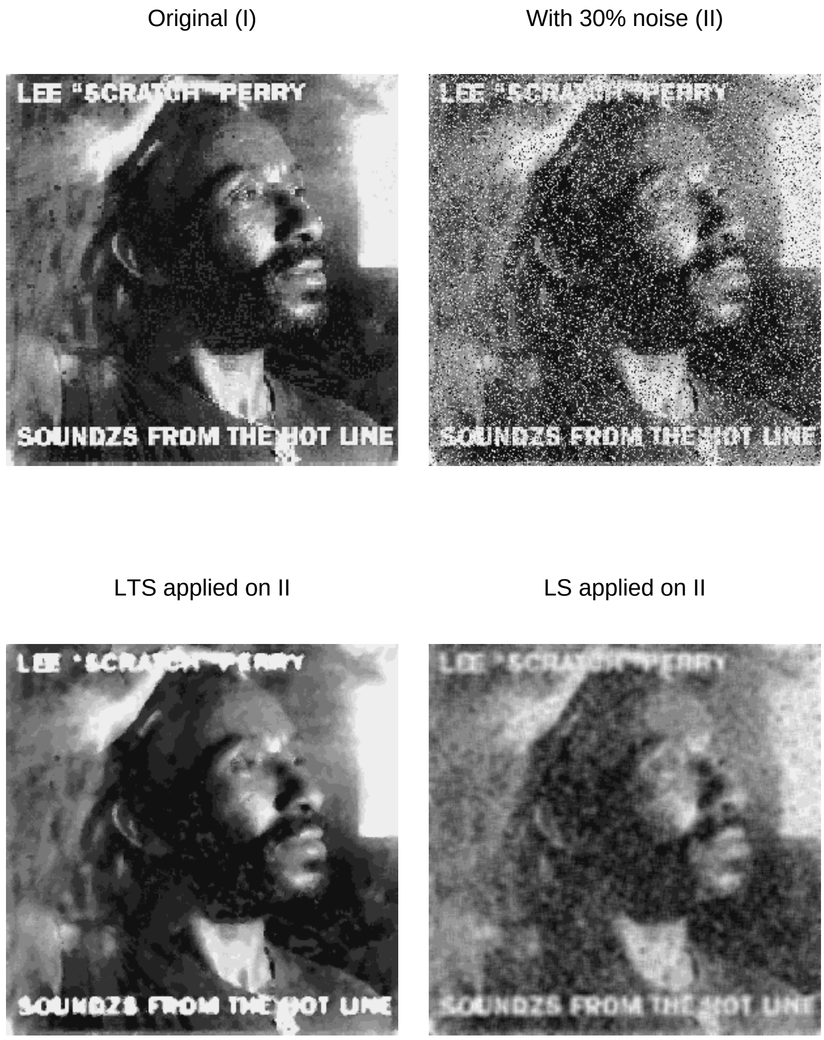

FIGURE 5. LTS and LS applied on the image of Lee Perry with 30% noise

corners are never preserved. This is due to the fact that the LTS and the LMS estimators are determined by the majority of the data.

Although the Cauchy estimator has highest breakdown point as well, it behaves not good. Its behavior is only a little bit better than that of the nonrobust LS estimator (see Figure 2). As the LS estimator, it is smoothing discontinuities. This bad behavior of the Cauchy estimator can have three reasons: Firstly, it could be that the estimator provided by S-PLUS is not exactly the Cauchy estimator. Secondly, the scale estimator within the Cauchy estimator could overestimate the scale. Thirdly, it could be that a high breakdown point is not the main property providing a good behavior in image analysis. Perhaps, it is more important for image representations that the estimator depends only on a majority of the data and is independent of a minority.

Acknowledgement. I am very grateful to my student Jan Flickschu who calculated the image representations.

References

Donoho, D.L. and Huber, P.J. (1983). The notion of breakdown point. In: P.J. Bickel, K.A. Doksum and J.L. Hodges, Jr., Eds., *A Festschrift for Erich L. Lehmann,* Wadsworth, Belmont, CA, 157-184.

He, X., Jurečková, J., Koenker, R. and Portnoy, S. (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica* 58, 1195–1214.

Meer, P., Mintz, D. and Rosenfeld, A. (1990). Least median of squares based robust analysis of image structure. *Proceedings of Image Understanding Workshop, DARPA*, 231-254.

Meer, P., Mintz, D., Rosenfeld, A. and Kim, D.Y. (1991). Robust regression methods in computer vision: A review. *International Journal of Computer Vision* **6**, 59-70.

Mizera, I. and Müller, C.H. (1996). Breakdown points and variation exponents of robust M-estimators in linear models. Preprint Nr. A-22-96, Fachbereich Mathematik und Informatik, Freie Universität Berlin.

Müller, Ch.H. (1995). Breakdown points for designed experiments. *J. Statist. Plann. Inference.* **45**, 413-427.

Müller, C.H. (1997). *Robust Planning and Analysis of Experiments.* Lecture Notes in Statistics **124**, Springer, New York.

Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection.* John Wiley, New York.

Smith, St.M. and Brady, J.M. (1995). SUSAN - A new approach to low level image processing. Technical Report TR95SMS1, Department of Clinical Neurology, Oxford University.