

Image Denoising: A Nonlinear Robust Statistical Approach

A. Ben Hamza and Hamid Krim, *Senior Member, IEEE*

Abstract—Nonlinear filtering techniques based on the theory of robust estimation are introduced. Some deterministic and asymptotic properties are derived. The proposed denoising methods are optimal over the Huber ϵ -contaminated normal neighborhood and are highly resistant to outliers. Experimental results showing a much improved performance of the proposed filters in the presence of Gaussian and heavy-tailed noise are analyzed and illustrated.

Index Terms— α -stable noise, asymptotic analysis, nonlinear filtering, robust estimation.

I. INTRODUCTION

LINEAR filtering techniques have been used in many image processing applications, and their popularity mainly stems from their mathematical simplicity and their efficiency in the presence of additive Gaussian noise. A mean filter is the optimal filter for Gaussian noise in the sense of mean square error. Linear filters, however, tend to blur sharp edges, destroy lines and other fine image details, fail to effectively remove heavy tailed noise, and perform poorly in the presence of signal-dependent noise. This led to a search for nonlinear filtering alternatives. The development of nonlinear median-based filters in recent years has resulted in remarkable results and has highlighted some new promising research avenues. On account of its simplicity, edge preservation property, and robustness to impulsive noise, the standard median filter remains the most popular for image processing applications [1]. The median filter, however, often tends to remove fine details in the image, such as thin lines and corners [1]. In recent years, a variety of median-type filters such as stack filters [2], multistage median [3], weighted median [4], rank conditioned rank selection [5], and relaxed median [6] have been developed to overcome this drawback.

The output of the relaxed median filter with parameters ℓ and u is determined by comparing lower and upper order statistics to the center sample in the filter window [6]. Its filtering operation is controlled by its parameters ℓ and u , which provide one with an ability to tradeoff between noise suppression and detail preservation. It has been shown that this type of filter provides improvement over other median-based filters in additive Gaussian white and heavy-tailed noise removal and pre-

serves details better than the standard median filter [6]. Roughly speaking, median-based filters are selective filters, that is, the output is always one of the input samples, and therefore, their use is more limited to impulsive noise removal.

A variety of models have been sources in modeling impulsive noise including the Laplacian model, whose distribution has heavier tails than the Gaussian. Examples of impulsive noise include atmospheric noise, cellular communication, underwater acoustics, and moving traffic. Recently, it has been shown that α -stable ($0 < \alpha \leq 2$) distributions can approximate impulsive noise more accurately than other models [7]. The parameter α controls the degree of impulsiveness (heaviness of the tails), and the impulsiveness increases as α decreases. The Gaussian ($\alpha = 2$) and the Cauchy ($\alpha = 1$) distributions are the only symmetric α -stable distributions that have closed-form probability density functions. The two most important properties of α -stable distributions are the *stability property* and the *generalized central limit theorem* [7].

The LogCauchy filter proposed in [8], as the maximum likelihood of the location parameter for the Cauchy ($\alpha = 1$) distribution, has been shown to be efficient in removing highly impulsive noise.

It is also known that in the presence of only Gaussian noise, the efficiency of a median filter leaves room for much improvement relative to that of a mean filter. This led to a number of other proposed nonlinear schemes to attain a balance between the two. Among these proposed filters, figure Wilcoxon, Hodges-Lehmann [1], and α -trimmed mean [9].

Other approaches of wavelet-based denoising have generally relied on the assumption on Gaussian noise [10] and are therefore sensitive to outliers, i.e., to noise distributions whose tails are heavier than the Gaussian distribution, such as Laplacian and α -stable distributions [7]. For ϵ -contaminated Gaussian distributions of the wavelet coefficients, Krim and Schick [11] derive a robust estimator of the wavelet coefficients for wavelet shrinkage and based on minimax description length.

In the next section, we provide a brief review of the Huber minimax approach, some basic sliding window filters, and symmetric α -stable ($S\alpha S$) distributions. In Section III, a competitive filtering scheme called the *mean-median* filter is proposed, and its asymptotic analysis is performed. In Section IV, we propose an efficient class of nonlinear filters called *mean-relaxed median* filters, which are defined through a convex combination of the mean and relaxed median filters. Section V is devoted to another class of nonlinear denoising techniques called *Mean-LogCauchy* filters. Finally, in Section VI, we provide experimental results to show a much-improved performance of the proposed filters and to substantiate our claims in this paper.

Manuscript received September 18, 2000; revised September 25, 2001. This work was supported by AFOSR Grant F49620-98-1-0190, by ONR-MURI Grant JHU-72798-S2, and by the North Carolina State University School of Engineering. The associate editor coordinating the review of this paper and approving it for publication was Dr. Ali H. Sayed.

The authors are with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695 USA (e-mail: abhamza@eos.ncsu.edu; ahk@eos.ncsu.edu).

Publisher Item Identifier S 1053-587X(01)10496-4.

II. BACKGROUND

Consider the additive noise model

$$X_{\mathbf{i}} = S_{\mathbf{i}} + V_{\mathbf{i}}, \quad \mathbf{i} \in \mathbb{Z}^m \quad (1)$$

where $\{S_{\mathbf{i}}\}$ is a discrete m -dimensional deterministic sequence corrupted by the zero-mean noise sequence $\{V_{\mathbf{i}}\}$, and $\{X_{\mathbf{i}}\}$ is the observed sequence. The objective is to estimate the sequence $S_{\mathbf{i}}$ based on a filtering output $Y_{\mathbf{i}} = \mathcal{F}(X_{\mathbf{i}})$, where \mathcal{F} is a filtering operator.

Here, we assume that the noise probability distribution is a scaled version of a *known* member of the family \mathcal{P}_ϵ of ϵ -contaminated normal distributions proposed by Huber [12], [13]

$$\mathcal{P}_\epsilon = \{(1 - \epsilon)\Phi + \epsilon H : H \in \mathcal{S}\}$$

where Φ is the standard normal distribution, \mathcal{S} is the set of all probability distributions symmetric with respect to the origin (i.e., such that $H(-x) = 1 - H(x)$), and $\epsilon \in [0, 1]$ is the known fraction of ‘‘contamination.’’ It is worth noting that under certain conditions and assuming that the distribution of the observations is an *unknown* member of the family \mathcal{P}_ϵ , Huber found the least-favorable distribution in the set \mathcal{P}_ϵ to have exponential tails in the no-process noise case and is simply the one minimizing the Fisher information [13]. The presence of outliers in a nominally normal sample can be modeled here by a distribution H with tails that are heavier than normal. Note that symmetry ensures the unbiasedness of the maximum likelihood estimator, making the expression for its asymptotic variance considerably simpler. Although this restriction obviously precludes cases where outliers are grouped on one side of the mean of the nominal (‘‘underlying’’) distribution, the model is general enough to represent many realistic situations. The asymmetric case has been studied in [14] and [15].

Krim and Schick [11] proposed a robust wavelet thresholding technique based on the minimax description length (MMDL) principle, determining the least favorable distribution in a ϵ -contaminated normal family as the member that maximizes the entropy. The MMDL approach results in a thresholding scheme that is resistant to heavy-tailed noise.

Let W be a sliding window of size $2N + 1$. Define $W_{\mathbf{i}} = \{X_{\mathbf{i}+\mathbf{r}} : \mathbf{r} \in W\}$ to be the window data sequence centered at location \mathbf{i} .

The output of the mean filter is given by

$$Y_{\mathbf{i}} = \overline{W}_{\mathbf{i}} = \arg \min_{\theta} \sum_{\mathbf{r} \in W} (X_{\mathbf{i}+\mathbf{r}} - \theta)^2 \quad (2)$$

where $\overline{W}_{\mathbf{i}}$ is the sample mean of $W_{\mathbf{i}}$, and θ is the estimation parameter.

Denote by $[W_{\mathbf{i}}]_{(k)}$ the k th-order statistic [6] of the samples in $W_{\mathbf{i}}$, that is

$$[W_{\mathbf{i}}]_{(1)} \leq [W_{\mathbf{i}}]_{(2)} \leq \dots \leq [W_{\mathbf{i}}]_{(2N+1)}.$$

The output of the standard median filter (SM) is given by

$$Y_{\mathbf{i}} = [W_{\mathbf{i}}]_{(N+1)} = \arg \min_{\theta} \sum_{\mathbf{r} \in W} |X_{\mathbf{i}+\mathbf{r}} - \theta|. \quad (3)$$

Such estimators are well founded and well known for a Gaussian and Laplacian distributions. Note that the mean and median filters are the maximum likelihood estimators of the location parameter for the Gaussian and Laplacian distributions, respectively.

The general class of α -stable distributions has also been shown to accurately model heavy-tailed noise [7]. A symmetric α -stable ($S\alpha S$) random variable is, however, only described by its characteristic function

$$\varphi(t) = \exp(j\theta t - \gamma|t|^\alpha)$$

where $j \in \mathbb{C}$ is the imaginary unit, $\theta \in \mathbb{R}$ is the location parameter (centrality), $\gamma \in \mathbb{R}$ is the dispersion of the distribution and $\alpha \in (0, 2]$, which controls the heaviness of the tails, is the characteristic exponent [7].

When $\alpha \in (0, 2)$, an $S\alpha S$ random variable has infinite variance, and the Cauchy ($\alpha = 1$) is the only distribution that has a closed form for the probability density function. This is, in fact, useful when using the principle of maximum likelihood estimation.

The LogCauchy (LC_γ) filter [8] is the maximum log-likelihood estimator of the location parameter for a Cauchy density and yields the following

$$Y_{\mathbf{i}} = LC_\gamma(W_{\mathbf{i}}) = \arg \min_{\theta} \sum_{\mathbf{r} \in W} \log(\gamma^2 + (X_{\mathbf{i}+\mathbf{r}} - \theta)^2) \quad (4)$$

where γ is the dispersion, and θ is the estimation parameter.

In the next section, a competitive nonlinear filter inspired by Huber’s approach to robust estimation is introduced, and its asymptotic variance is derived.

III. MEAN-MEDIAN FILTER

Robust estimation answers the need raised by the common situation where the distribution function is in fact *not* precisely known. In this case, a reasonable approach would be to assume that the density is a member of some set, or some family of parametric families, and to choose the *best* estimate for the *least favorable* member of that set. Huber [12] proposed an ϵ -contaminated normal set \mathcal{P}_ϵ and found that the least favorable distribution in \mathcal{P}_ϵ that maximizes the asymptotic variance (or, equivalently, minimizes the Fisher information) is Gaussian in the center and Laplacian in the tails and switches from one to the other at a point whose value depends on the fraction of contamination ϵ , larger fractions corresponding to smaller switching points, and vice versa.

From (2) and (3), it can easily be seen that the mean filter is optimal for Gaussian noise in the sense of mean square error, whereas the standard median filter for Laplacian noise in the sense of mean absolute error. Assume that the noise probability distribution P is a scaled version of a member of \mathcal{P}_ϵ , such that $P = (1 - \epsilon)G + \epsilon L$, where G is Gaussian $\mathcal{N}(0, \sigma_G^2)$ with variance σ_G^2 , and L is Laplacian (or double-exponential) $\mathcal{L}(0, \sigma_L^2)$ with variance σ_L^2 (clearly, $L \in \mathcal{S}$). The most commonly used form in modeling outliers for detection and robustness studies is the two-component mixture, where both distributions are zero mean, but one has greater variance than the other. Although the tails of the normal distribution are relatively light, this model is

the basis of a number of robust estimators in the literature. Our above assumption for the noise to be ϵ -contaminated Gaussian and Laplacian distributed is mainly due to the fact that heavier tails than the Gaussian mixture are provided by the Laplace distribution, which is used as a contaminant of the Gaussian distribution. It is worth noting that as shown above, Huber found the least favorable member of \mathcal{P}_ϵ to have exponential tails (in the no process noise case).

From the above discussion, a convex combination of the mean and the median filters can be defined as follows.

Definition 1: The output of the Mean-Median (MEM) filter is given by

$$Y_{\mathbf{i}} = (1 - \lambda)\overline{W}_{\mathbf{i}} + \lambda[W_{\mathbf{i}}]_{(N+1)}, \quad \lambda \in [0, 1].$$

As a suitable performance measure for a robust estimator, Huber suggests its asymptotic variance since the sample variance is strongly dependent on the tails of the distribution. Indeed, for any estimator whose value is always contained within the convex hull of the observations, the supremum of its actual variance is infinite. For this reason, the performance of the mean-median filter is carried out using its asymptotic variance.

The asymptotic variance $V(T, F)$ of an estimator T at the distribution F is given by

$$V(T, F) = \int IF(x; T, F)^2 dF(x) \tag{5}$$

where $IF(x; T, F)$ is the influence function of T at F , which is defined as

$$IF(x; T, F) = \lim_{t \rightarrow 0} \frac{T((1-t)F + t\Delta_x) - T(F)}{t}$$

at all points x where the limit exists, and Δ_x stands for delta distribution function, i.e., with unit mass at x . The influence function gives the effect of an infinitesimal perturbation to the data at the point x .

It can be shown that the influence function of the mean and the standard median filters are given by [13]

$$IF(x; \overline{W}_{\mathbf{i}}, F_\theta) = x - \theta$$

and

$$IF(x; [W_{\mathbf{i}}]_{(N+1)}, F_\theta) = \frac{\text{sign}(x - \theta)}{2f(\theta)}$$

where $F_\theta = F(x - \theta)$ is the common distribution function of the input, and f is the corresponding density function. It follows that the influence function of the MEM filter is given by

$$IF(x; \text{MEM}, F_\theta) = (1 - \lambda)(x - \theta) + \lambda \frac{\text{sign}(x - \theta)}{2f(\theta)}. \tag{6}$$

Using (5) and (6), we obtain the following result.

Proposition 1: The asymptotic variance $V(\text{MEM}, F_\theta)$ of the MEM filter at the distribution F_θ is given by

$$V(\text{MEM}, F_\theta) = (1 - \lambda)^2 \mu_2 + \frac{\lambda^2}{4f(\theta)^2} + \lambda(1 - \lambda) \frac{\mu_1}{f(\theta)} \tag{7}$$

where $\mu_k = E|X - \theta|^k, k = 1, 2$ are the central moments.

Remark: While the independence assumption of the filter input simplifies the tractability of the problem, it is not strictly valid.

Choosing the asymptotic variance as performance measure, it is necessary to obtain the minimum attainable asymptotic variance. The filter attaining that minimum asymptotic variance (i.e., minimizing over λ) will then provide the best filtering performance.

Corollary 1: The minimum value of $V(\text{MEM}, F_\theta)$ is attained at λ_{\min} which is given by

$$\lambda_{\min} = \left(\mu_2 - \frac{\mu_1}{2f(\theta)} \right) / \left(\mu_2 + \frac{1}{4f(\theta)^2} - \frac{\mu_1}{f(\theta)} \right). \tag{8}$$

Proof: Denote by $\mathcal{V}(\lambda) = V(\text{MEM}, F_\theta)$. Using (7), the asymptotic variance of the MEM filter can be written as

$$\mathcal{V}(\lambda) = \left(\mu_2 + \frac{1}{4f(\theta)^2} - \frac{\mu_1}{f(\theta)} \right) \lambda^2 + \left(\frac{\mu_1}{f(\theta)} - 2\mu_2 \right) \lambda + \mu_2.$$

Differentiating $\mathcal{V}(\lambda)$ with respect to λ and solving $\mathcal{V}'(\lambda) = 0$, we obtain a critical point λ_0 given by

$$\lambda_0 = \left(\mu_2 - \frac{\mu_1}{2f(\theta)} \right) / \left(\mu_2 + \frac{1}{4f(\theta)^2} - \frac{\mu_1}{f(\theta)} \right).$$

The second derivative of $\mathcal{V}(\lambda)$ is positive

$$\mathcal{V}''(\lambda) = 2 \left(\left(\sqrt{\mu_2} - \frac{1}{2f(\theta)} \right)^2 + \frac{1}{f(\theta)} (\sqrt{\mu_2} - \mu_1) \right) > 0$$

since $\sqrt{\mu_2} \geq \mu_1$, by the Cauchy-Schwartz inequality. Then, λ_0 is the global minimum of \mathcal{V} , and the proof is completed. ■

In the following example, the minimum values λ_{\min} for some common probability distributions are derived.

Example: To see how successful the median filter is in improving the mean filter for heavy-tailed noise distributions, let us consider the asymptotic relative efficiency (ARE) of $[W_{\mathbf{i}}]_{(N+1)}$ with respect to $\overline{W}_{\mathbf{i}}$ for an arbitrary symmetric distribution F . The ARE is defined as a ratio between the asymptotic variances, that is

$$\text{ARE}([W_{\mathbf{i}}]_{(N+1)}, \overline{W}_{\mathbf{i}}) = \frac{V(\overline{W}_{\mathbf{i}}, F_\theta)}{V([W_{\mathbf{i}}]_{(N+1)}, F_\theta)}.$$

Suppose that the input is i.i.d. with variance $\sigma^2 < \infty$; then

$$\sqrt{2N+1}(\overline{W}_{\mathbf{i}} - \theta) \xrightarrow{D} \mathcal{N}(0, \sigma^2),$$

and

$$\sqrt{2N+1}([W_{\mathbf{i}}]_{(N+1)} - \theta) \xrightarrow{D} \mathcal{N}\left(0, \frac{1}{4f(\theta)^2}\right)$$

provided F_θ has a density f with $f(\theta) > 0$. It follows that

$$\text{ARE}([W_{\mathbf{i}}]_{(N+1)}, \overline{W}_{\mathbf{i}}) = 4\sigma^2 f(\theta)^2.$$

If the input is i.i.d. Gaussian $\mathcal{N}(\mu, \sigma^2)$, then because $f(\mu) = 1/(\sqrt{2\pi}\sigma)$, we have $\text{ARE} = 2/\pi \approx 0.637$. Therefore, the mean filter performs better than the median filter in removing

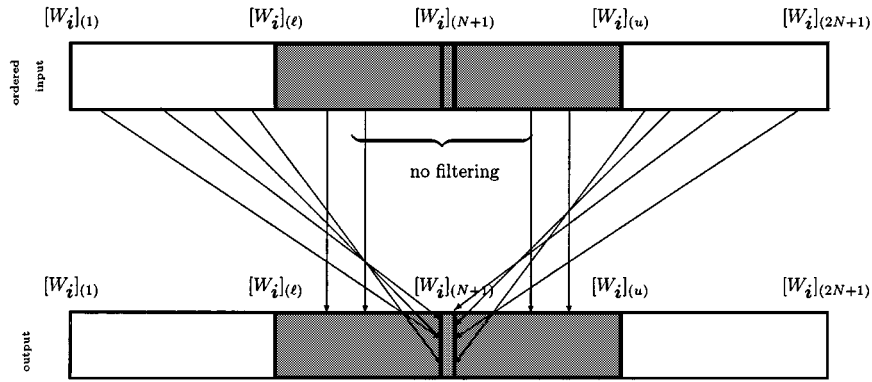


Fig. 1. Relaxed median filtering operation.

Gaussian noise. Using (8), it is easy to check that the minimum value of $V(\text{MEM}, F_\theta)$ is attained at $\lambda_{\min} = 2/(\pi + 2)$.

Similarly, if the input is Laplacian $\mathcal{L}(\mu, \sigma^2)$ with mean μ and variance σ^2 , then $\text{ARE} = 2$. Thus, the median filter has better performance than the mean filter in the presence of Laplacian noise. The minimum value of $V(\text{MEM}, F_\theta)$ is attained at $\lambda_{\min} = 2/3$.

If we stretch the tail of the distribution, thereby increasing σ^2 , the efficiency will increase and tend to ∞ as $\sigma^2 \rightarrow \infty$. A limiting case is the Cauchy distribution ($S\alpha S$ distribution with $\alpha = 1$) in which $\sigma^2 = \infty$ so that $\text{ARE} = \infty$.

A generalization of the median filters is the class of L-filters [16]. Its output is given by

$$Y_{\mathbf{i}} = \sum_{k=1}^{2N+1} a_k [W_{\mathbf{i}}]_{(k)}$$

where a_k are the filter coefficients. It is easy to show that the MEM filter is a special case of L-filters with corresponding weights

$$a_k = \begin{cases} \frac{1+2N\lambda}{2N+1}, & \text{if } k = N+1 \\ \frac{1-\lambda}{2N+1}, & \text{otherwise.} \end{cases}$$

Since the relaxed median filter provides improvement over the standard median filter in preserving details and removes noise better than other median-based filters, a novel class of nonlinear filters as a convex combination of the mean and the relaxed median is defined in the next section.

IV. MEAN-RELAXED MEDIAN FILTERS

Definition 2: The output of the relaxed median ($\text{RM}_{\ell u}$) filter with parameters ℓ and u is given by

$$Y_{\mathbf{i}} = \text{RM}_{\ell u}(W_{\mathbf{i}}) = \begin{cases} X_{\mathbf{i}}, & \text{if } X_{\mathbf{i}} \in [[W_{\mathbf{i}}]_{(l)}, [W_{\mathbf{i}}]_{(u)}] \\ [W_{\mathbf{i}}]_{(N+1)}, & \text{otherwise} \end{cases} \quad (9)$$

where ℓ and u are such that $1 \leq \ell \leq N+1 \leq u \leq 2N+1$. Note that when $\ell = 1$ and $u = 2N+1$, the $\text{RM}_{\ell u}$ filter becomes the identity filter (no filtering), and when $\ell = u = N+1$, the output of the $\text{RM}_{\ell u}$ is simply the median. Fig. 1 illustrates the structure of the relaxed median filter.

It is worth noting that the symmetric relaxed median filter is also called rank conditioned median filter [5]. In [6], the output

distribution of the relaxed median filter is given in a more generalized form, and its statistical properties for details preservation are also presented and illustrated.

A convex combination of the mean and relaxed median filters gives rise to a competitive class of nonlinear filters defined as follows.

Definition 3: The output of the mean-relaxed median ($\text{MRM}_{\ell u}$) filters with parameters ℓ and u is given by

$$Y_{\mathbf{i}} = \text{MRM}_{\ell u}(W_{\mathbf{i}}) = (1-\lambda)\overline{W}_{\mathbf{i}} + \lambda \text{RM}_{\ell u}(W_{\mathbf{i}}) \quad (10)$$

where $\lambda \in [0, 1]$, and $1 \leq \ell \leq N+1 \leq u \leq 2N+1$.

The parameters ℓ, u and λ allow the $\text{MRM}_{\ell u}$ filter to have a variety of characteristics. Note that when $\ell = 1$ and $u = 2N+1$, the output of the mean-relaxed median filter becomes

$$Y_{\mathbf{i}} = (1-\lambda)\overline{W}_{\mathbf{i}} + \lambda X_{\mathbf{i}}$$

and if we choose $\lambda = \sigma_S^2 / (\sigma_S^2 + \sigma_V^2)$, where σ_S^2 is the sample variance of the original image and σ_V^2 is the noise variance, then the mean-relaxed median filter is simply the local linear minimum mean square error (LLMMSE) filter [17].

As two important special cases, when $\lambda = 1$, we obtain the relaxed median filter, and when $\ell = u$, the $\text{MRM}_{\ell u}$ becomes the mean-median filter.

An important statistical property of the relaxed median filter is the probability that the center sample of the windowed sequence be the output, in other words, the probability that a sample remains unchanged by the relaxed median filter. This probability gives some information about the filter capacity to preserve details. Additional statistical properties are studied in detail in [6]. The following result follows directly from the definition of relaxed median filters.

Proposition 2: Let the input of the $\text{RM}_{\ell u}$ filter be i.i.d. with a continuous distribution function; then

$$\Pr[Y_{\mathbf{i}} = X_{\mathbf{i}}] = \frac{u - \ell + 1}{2N + 1}$$

for all ℓ and u such that $1 \leq \ell \leq N+1 \leq u \leq 2N+1$.

Note that as ℓ decreases, the probability that the relaxed median filter outputs a center sample increases, and therefore, the relaxed median filter tends to preserve more details.

To show the improvement of the mean-relaxed median over the mean-median in detail preservation, some deterministic properties of relaxed median filters such as roots and constant neighborhoods are derived in the next subsection. Statistical properties for detail preservation can be found in [6].

A. Properties of Relaxed Median Filters

In this subsection, some deterministic properties of relaxed median filters are presented. First, it is easy to show that relaxed median filters are translation and scale invariant, that is

$$\text{RM}_{\ell u}(sW_{\mathbf{i}} + t) = s\text{RM}_{\ell u}(W_{\mathbf{i}}) + t, \quad \forall s, t \in \mathbb{R}.$$

For the sake of simplicity, we consider a symmetric relaxed median filter, that is, the parameters ℓ and u are symmetric ($u = 2N + 2 - \ell$). This restriction is appropriate when the signal and noise are symmetric and to ensure the unbiasedness of the filter. Therefore, the symmetric relaxed median $\text{RM}(2N + 1; \ell)$ filter is completely specified by two parameters: the window width and the lower bound ℓ . The deterministic behavior of the relaxed median filter can be analyzed by considering its effect on arbitrary sequences. The following result gives a necessary and sufficient condition for an arbitrary signal to be a root (fixed point) of the relaxed median filter.

Proposition 3: A signal is a root of $\text{RM}(2N + 1; \ell)$ if and only if there are at least ℓ samples with the same value as the center sample of the filter window.

Proof: According to (9), a signal is a root of the relaxed median filter if and only if $[W_{\mathbf{i}}]_{(\ell)} \leq X_{\mathbf{i}} \leq [W_{\mathbf{i}}]_{(2N+2-\ell)}$, that is, if and only if there are at least ℓ samples with the same value as $X_{\mathbf{i}}$. In other words, the filter window contains no less than ℓ samples with the same value as the center sample. \square

When $\ell = N + 1$, a relaxed median filter is just a standard median filter, in which case, it is well known that only signals with constant neighborhoods (minimum length $N + 1$) and edges (monotonic regions between two constant neighborhoods of different values) are roots of the standard median filter.

According to Proposition 3, it is straightforward to prove the following result.

Proposition 4: The minimum length of a constant neighborhood of $\text{RM}(2N + 1; \ell)$ is ℓ .

Note that as ℓ decreases, the minimum length of a constant neighborhood decreases, allowing finer details to be preserved.

The following result illustrates the relationship among root signal sets of relaxed median filters with the same window width and different lower bounds.

Proposition 5: If R_{2N+1}^{ℓ} denotes the root set of $\text{RM}(2N + 1; \ell)$, then

$$\cdots \subseteq R_{2N+1}^{\ell} \subseteq R_{2N+1}^{\ell-1} \subseteq \cdots \subseteq S$$

where S is the set of all signals.

Proof: A root signal of R_{2N+1}^{ℓ} is also a root of $R_{2N+1}^{\ell-1}$. Indeed, according to (9) and to the definition of order statistics, $[W_{\mathbf{i}}]_{(\ell)} \leq X_{\mathbf{i}} \leq [W_{\mathbf{i}}]_{(2N+2-\ell)}$ implies $[W_{\mathbf{i}}]_{(\ell-1)} \leq X_{\mathbf{i}} \leq [W_{\mathbf{i}}]_{(2N+3-\ell)}$. The result follows by induction.

The LogCauchy filter has been shown to outperform the standard median filter in removing highly α -stable noise [8]; then,

the MEM filter can be improved, replacing the median by the LogCauchy, and therefore, a new class of nonlinear filters is derived and will be defined in the next section.

V. MEAN-LOGCAUCHY FILTERS

Analogously to Section III, we now assume that the noise probability distribution P is a scaled version of a member of \mathcal{P}_{ϵ} such that $P = (1 - \epsilon)G + \epsilon S$, where G is Gaussian $\mathcal{N}(0, \sigma_G^2)$, and S is $S_{\alpha}S$ with location parameter θ and dispersion γ_S . The parameter α controls how impulsive the distribution is.

Suppose that G and S are the cumulative distribution functions of two independent random variables X_G and X_S , respectively. Then, it is easy to show that the characteristic function φ_{ϵ} of the random variable $(1 - \epsilon)X_G + \epsilon X_S$ is given by

$$\varphi_{\epsilon}(t) = \exp\left(j\epsilon\theta t - (1 - \epsilon)^2 \frac{\sigma_G^2}{2} t^2 - \epsilon^{\alpha} \gamma_S |t|^{\alpha}\right), \quad \epsilon \in [0, 1].$$

For $\alpha \in (1, 2]$, all $S_{\alpha}S$ random variables have finite mean given by their location parameter θ . Moreover, it is shown in [18] that an $S_{\alpha}S$ distribution with zero mean can be approximated by a finite-Gaussian mixture. Assuming that S is zero mean $S_{\alpha}S$ ($1 < \alpha \leq 2$), then $P = (1 - \epsilon)G + \epsilon S$ can be approximated by a finite-Gaussian mixture, and hence, the noise model (1) becomes an ϵ -contaminated Gaussian mixture noise model.

For $\alpha \in (0, 1]$, all $S_{\alpha}S$ random variables have a median, and the only $S_{\alpha}S$ distribution that has a closed-form probability density function is Cauchy distribution ($\alpha = 1$); thus, the maximum log-likelihood principle can be applied to derive (4).

A convex combination of the mean and the LogCauchy filters can then be defined as follows.

Definition 4: The output of the Mean-LogCauchy (MLC_{γ}) filter with parameter γ is given by

$$Y_{\mathbf{i}} = \text{MLC}_{\gamma}(W_{\mathbf{i}}) = (1 - \lambda)\overline{W}_{\mathbf{i}} + \lambda \text{LC}_{\gamma}(W_{\mathbf{i}}) \quad (11)$$

where $\lambda \in [0, 1]$, and γ is the dispersion of a Cauchy distribution.

The output of the LogCauchy filter is defined as a solution of the following maximum log-likelihood estimation problem

$$\begin{aligned} \hat{\theta}_{\mathbf{i}} &= \arg \max_{\theta} \ell_{\gamma}(\theta W_{\mathbf{i}}) \\ &= \arg \max_{\theta} \log \prod_{\mathbf{r} \in W} \frac{\gamma}{\pi} \left(\frac{1}{\gamma^2 + (X_{\mathbf{i}+\mathbf{r}} - \theta)^2} \right) \end{aligned} \quad (12)$$

where $\ell_{\gamma}(\theta; W_{\mathbf{i}})$ is the log-likelihood function of a Cauchy distribution $\mathcal{C}(\gamma, \theta)$.

It is clear that for a given γ , solving (12) is equivalent to minimizing the function $\rho_{\gamma}(\theta; W_{\mathbf{i}})$ given by

$$\rho_{\gamma}(\theta; W_{\mathbf{i}}) = \prod_{\mathbf{r} \in W} (\gamma^2 + (X_{\mathbf{i}+\mathbf{r}} - \theta)^2) \quad (13)$$

as well as to solving the problem (4) since the $\log(\cdot)$ function is strictly monotone. Thus, the minimum of (4) is attained at the same place as that of $\rho_{\gamma}(\theta; W_{\mathbf{i}})$. This is very important because $\rho_{\gamma}(\theta; W_{\mathbf{i}})$ is a polynomial of degree $2(2N + 1)$

in θ , and its characteristics can then be obtained easily. It is shown in [19] that $\rho_\gamma(\theta; W_{\mathbf{i}})$ is a convex function of θ if $\gamma \geq [W_{\mathbf{i}}]_{(2N+1)} - [W_{\mathbf{i}}]_{(1)}$ and, therefore, has a unique minimum $\theta_0 \in [[W_{\mathbf{i}}]_{(1)}, [W_{\mathbf{i}}]_{(2N+1)}]$. At $\gamma = 0$, the function $\rho_\gamma(\theta; W_{\mathbf{i}})$ has distinct minima at all the points $X_{\mathbf{i}+\mathbf{r}}$. If γ is increased, the number of distinct minima decreases. After a certain limit of γ , there is only a unique minimum. Obviously, γ cannot be arbitrarily large as it has to satisfy (12).

It has been shown in [20] that the myriad (LogCauchy) filter possesses this important property; it converges to the mean filter when $\gamma \rightarrow \infty$, i.e.,

$$\text{LC}_\gamma(W_{\mathbf{i}}) \rightarrow \overline{W_{\mathbf{i}}} \quad \text{as } \gamma \rightarrow \infty. \quad (14)$$

Asymptotically, the tuning parameter γ transforms a nonlinear filter to a linear one. Therefore, γ is also called the *linearity parameter* of the myriad filter [20].

Hence, the Mean-LogCauchy filter becomes the mean filter when $\gamma \rightarrow \infty$. To this end, the following result gives an alternative and a simple proof for (14).

Proposition 6: When $\gamma \rightarrow \infty$, the Mean-LogCauchy filter becomes the mean filter, i.e.,

$$\text{MLC}_\gamma(W_{\mathbf{i}}) \rightarrow \overline{W_{\mathbf{i}}} \quad \text{as } \gamma \rightarrow \infty.$$

Proof: Using basic properties of the arg min function, the output of the LogCauchy filter can be expressed as

$$\begin{aligned} \text{LC}_\gamma(W_{\mathbf{i}}) &= \arg \min_{\theta} \sum_{\mathbf{r} \in W} \log(\gamma^2 + (X_{\mathbf{i}+\mathbf{r}} - \theta)^2) \\ &= \arg \min_{\theta} \sum_{\mathbf{r} \in W} \gamma^2 \log \left(1 + \frac{(X_{\mathbf{i}+\mathbf{r}} - \theta)^2}{\gamma^2} \right) \\ &= \arg \min_{\theta} \sum_{\mathbf{r} \in W} \log \left(1 + \frac{(X_{\mathbf{i}+\mathbf{r}} - \theta)^2}{\gamma^2} \right)^{\gamma^2}. \end{aligned}$$

Since

$$\lim_{\gamma \rightarrow \infty} \log \left(1 + \frac{(X_{\mathbf{i}+\mathbf{r}} - \theta)^2}{\gamma^2} \right)^{\gamma^2} = \exp\{(X_{\mathbf{i}+\mathbf{r}} - \theta)^2\}$$

and the exponential function $\exp\{\cdot\}$ is monotonically increasing, it follows that

$$\text{LC}_\gamma(W_{\mathbf{i}}) \rightarrow \arg \min_{\theta} \sum_{\mathbf{r} \in W} (X_{\mathbf{i}+\mathbf{r}} - \theta)^2 \quad \text{as } \gamma \rightarrow \infty.$$

This concludes the proof using (11) and (14). ■

VI. SIMULATION RESULTS

This section presents simulation results where the proposed filters are applied to enhance images corrupted by impulsive noise as well as ϵ -contaminated mixed noise. The performance of a filter clearly depends on the filter type and its sliding window size, the properties of signals/images, and the characteristics of the noise. The choice of criteria by which to measure the performance of a filter presents certain difficulties. In particular, it is clear that a global performance measure such as the mean square error only gives a partial picture of reality; for instance, one filter may do very well at the nominal model but badly at an outlier, whereas another may do poorly at the

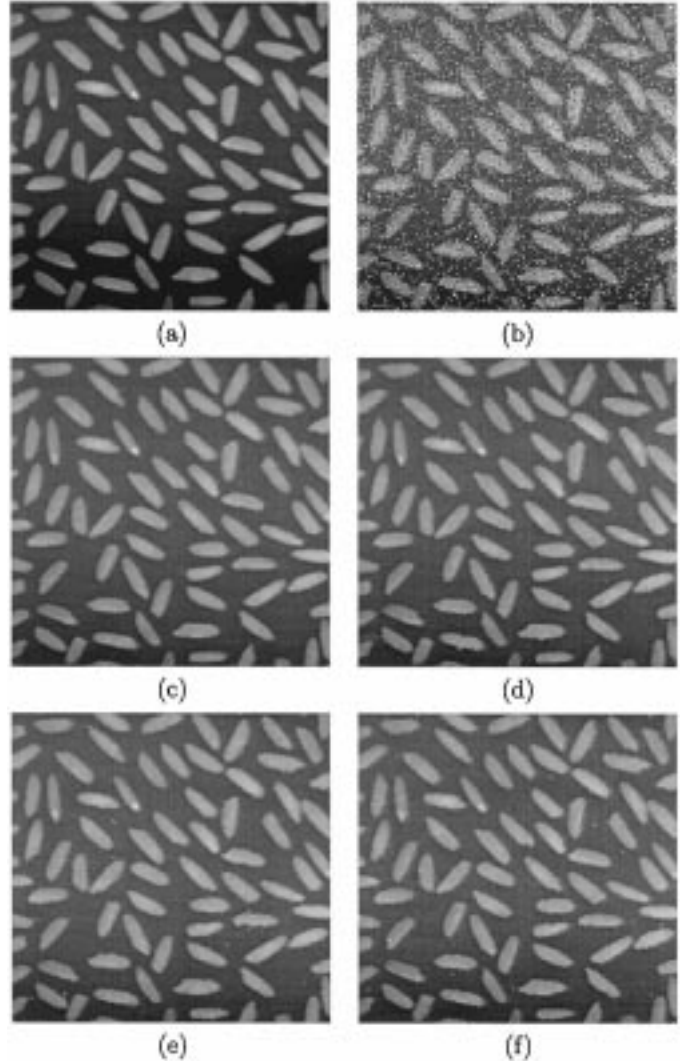


Fig. 2. Filtering results for impulsive noise using a 3×3 square window. (a) Original image. (b) Noisy image ("Salt and Pepper noise" 10%). (c) Output of the SM filter. (d) Output of the $\text{RM}_{4,6}$ filter. (e) Output of the $\text{CWM}(\omega_c = 3)$ filter. (f) Output of the tri-state median filter.

nominal model but well at an outlier, and yet, the two could have the same mean square value.

Another important performance measure in the mean absolute error obviously tends to downplay the influence of large errors, compared with mean square error, precisely in the presence of heavy-tailed noise.

To assess the performance of the proposed filters, mean square error (MSE) between the filtered and the original image is evaluated to quantitatively compare the performance of these proposed techniques with other filtering schemes.

A. Relaxed Median Performance in Impulsive Noise

In order to evaluate the performance of relaxed median filters in the presence of heavy-tailed noise, the image shown in Fig. 2(a) has been corrupted by "Salt and Pepper noise" ($p = 0.1$). Fig. 2 displays the results of filtering the noisy image shown in Fig. 2(a) by a standard median (SM) filter, a relaxed median filter ($\text{RM}_{4,6}$), a center weighted median (CWM) filter [4] with central weight $\omega_c = 3$, and a tri-state median filter [21]

TABLE I
MSEs RESULTS FOR IMPULSIVE NOISE

Filter type (3 × 3)	MSE		
	$p = 0.1$	$p = 0.15$	$p = 0.2$
SM	15.6750	28.9598	76.1020
RM _{4,6}	20.0122	85.4673	185.6687
CWM($\omega_c = 3$)	46.2378	89.1845	202.9764
Tri-State Median	52.9664	106.7337	204.7532

with central weight $\omega_c = 3$ and threshold $T = 20$. Qualitatively, we observe that the relaxed median filter RM_{4,6} is able to suppress impulsive noise while preserving important features in the image. Table I compares the resulting MSE computations.

To demonstrate the performance of the relaxed median filter in the presence of a noisy step edge, consider a widely used model for the step edge, which results from the blurring of the ideal step edge with Gaussian point spread function. A typical 2-D step edge is given by

$$\mathcal{E}(x, y) = \frac{a}{2} \left(1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) + b \right)$$

where a and b are constants, and $\operatorname{erf}(\cdot)$ is the error function.

An example of this edge is shown in Fig. 3(a). The α -stable ($\alpha = 0.8$) noisy step edge is shown in Fig. 3(b). The filtering results using the standard median, the relaxed median, the center weighted median, and the tri-state median are illustrated in Fig. 3. It is clear that the relaxed median filter preserves the edge well while removing α -stable noise.

B. Mean-Relaxed Median Performance in Mixed Noise

The Laplacian noise is somewhat heavier than the Gaussian noise. Moreover, the Laplace distribution is similar to Huber’s least favorable distribution [11] (for the no process noise case), at least in the tails. To demonstrate the application of mean-relaxed median filters to image denoising, qualitative and quantitative comparisons are performed to show the advantage of these filters over existing techniques. Fig. 4(b) shows a noisy image contaminated by ϵ -contaminated mixed Gaussian white noise $\mathcal{N}(0, 100)$ and Laplacian white noise $\mathcal{L}(0, 400)$. The fraction of contamination ϵ is chosen to be equal to λ in all the experiments. Table II summarizes the MSEs results obtained by applying the MEM, MRM_{4,6}, Wilcoxon, and Hodges-Lehmann filters to the noisy image with the results shown in Fig. 4.

Note that the MRM_{4,6} filter outperforms Wilcoxon and Hodges-Lehmann filters in reducing mixed noise, whereas the MEM filter achieves the best performance. Comparison of these images clearly indicates that the MRM_{4,6} filter preserves details well while removing mixed noise.

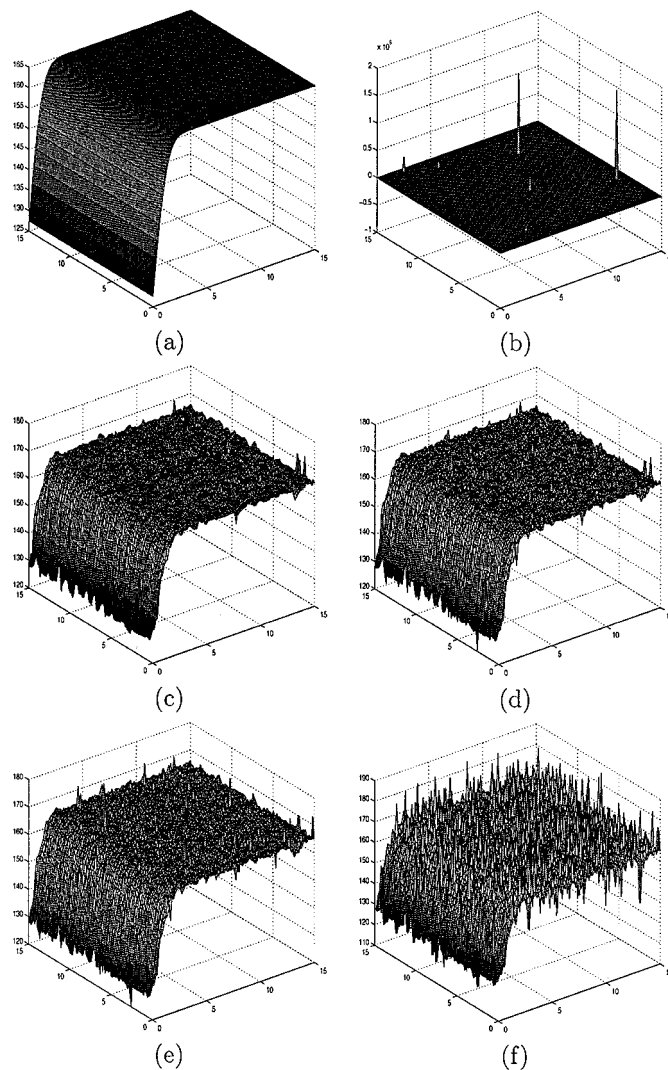


Fig. 3. Edge and filtering results for α -stable noise. (a) Step edge. (b) Noisy step edge (α -stable noise $\alpha = 0.8$). (c) Output of the SM filter. (d) Output of the RM_{4,6} filter. (e) Output of the CWM($\omega_c = 3$) filter. (f) Output of the tri-state median filter.

C. Mean-LogCauchy Performance in Mixed Noise

The scale-contaminated Gaussian and Laplace distributions are relatively light tailed. The $S\alpha S$ distributions are very heavy-tailed noise distributions. The Cauchy distribution is a member of this family ($\alpha = 1$), whose variance is infinite. To assess the performance of Mean-LogCauchy filters in mixed noise, the original image in Fig. 5(a) was contaminated by both Gaussian white noise $\mathcal{N}(0, 100)$ and α -stable noise $S\alpha S(\alpha = 0.5)$. The ϵ -contaminated mixed noise corrupted image is shown in Fig. 5(b). The visual comparison with other techniques is shown in Fig. 5.

Table III compares several filters quantitatively using the MSE criterion. The Mean-LogCauchy filter achieves the best performance. The RM_{4,6} outperforms Wilcoxon and Hodges-Lehmann in suppressing highly α -stable noise, that is, for $\alpha \in (0, 1]$. This substantiates our discussion about the parameter α in Section V because for $\alpha \in (1, 2]$, the noise

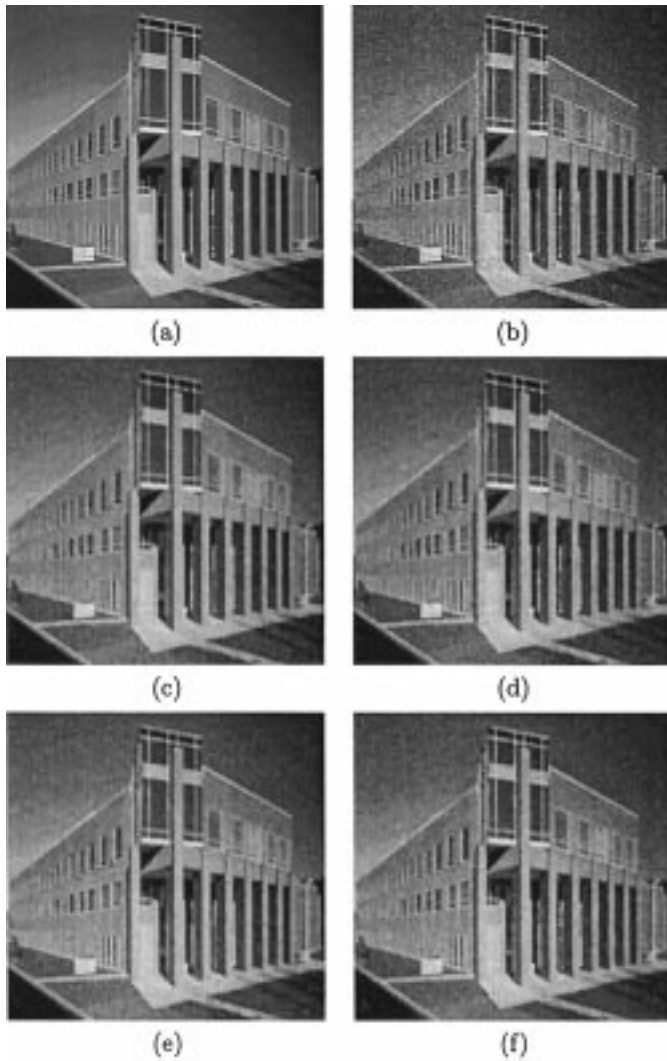


Fig. 4. Filtering results for ϵ -contaminated Gaussian and Laplacian noise using a 3×3 square window. (a) Original image. (b) ϵ -mixed noisy image: $\mathcal{N}(0, 100)$ and $\mathcal{L}(0, 400)$. (c) Output of the MEM filter $\lambda = 2/(2 + \pi)$. (d) Output of the $\text{MRM}_{4,6}$ filter $\lambda = 2/(2 + \pi)$. (e) Output of the Wilcoxon filter. (f) Output of the Hodges-Lehmann filter.

TABLE II
MSES RESULTS FOR ϵ -CONTAMINATED MIXED GAUSSIAN AND LAPLACIAN NOISE

Filter type (3×3)	MSE: $(1 - \epsilon)\mathcal{N}(0, 100) + \epsilon\mathcal{L}(0, \sigma_L^2)$		
	$\sigma_L^2 = 100$	$\sigma_L^2 = 200$	$\sigma_L^2 = 400$
MEM ($\lambda = \frac{2}{2+\pi}$)	184.4266	191.8701	207.0494
$\text{MRM}_{4,6}$ ($\lambda = \frac{2}{2+\pi}$)	185.4029	193.4539	209.4939
Wilcoxon	214.4146	216.6895	225.0940
Hodges-Lehmann	223.8270	226.6227	233.7648

distribution is approximately a Gaussian mixture, and therefore, the relaxed median filter is not very effective in removing Gaussian noise compared with Wilcoxon and Hodges-Lehmann filters.

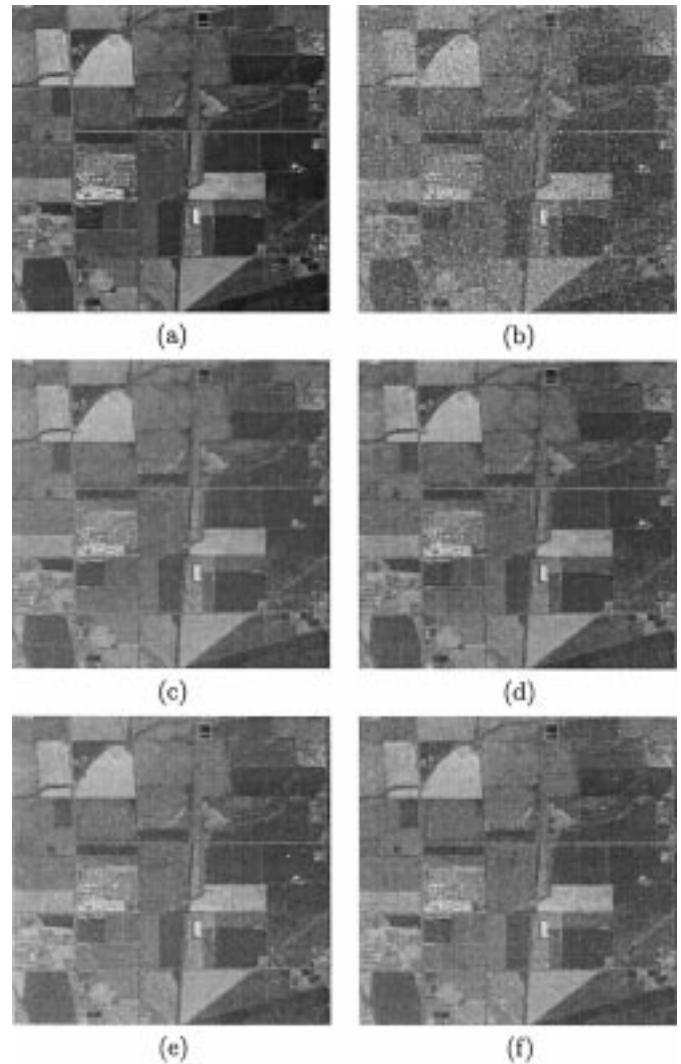


Fig. 5. Filtering results using a ϵ -contaminated Gaussian and α -stable noise 3×3 square window. (a) Original image. (b) ϵ -mixed noisy image: $\mathcal{N}(0, 100)$ and $S\alpha S$. (c) Output of the MLC filter $\lambda = 2/(2 + \pi)$. (d) Output of the $\text{RM}_{4,6}$ filter. (e) Output of the Wilcoxon filter. (f) Output of the Hodges-Lehmann filter.

TABLE III
MSES RESULTS FOR ϵ -CONTAMINATED MIXED GAUSSIAN AND α -STABLE NOISE

Filter type (3×3)	MSE: $(1 - \epsilon)\mathcal{N}(0, 100) + \epsilon(S\alpha S)$		
	$\alpha = 0.5$	$\alpha = 1.2$	$\alpha = 1.7$
MLC_{10} ($\lambda = \frac{2}{2+\pi}$)	65.2731	50.7154	49.0220
$\text{RM}_{4,6}$	68.8093	57.5971	56.6970
Wilcoxon	395.6731	53.3316	52.6859
Hodges-Lehmann	405.7872	54.2273	53.6443

D. Influence of the Parameters on the Proposed Filters

The high sensitivity of many specific filters to an accurate modeling of noise that is to be removed led us to investigate the proposed new techniques that include a number of filters whose optimality, when given a specific noise distribution, is attained

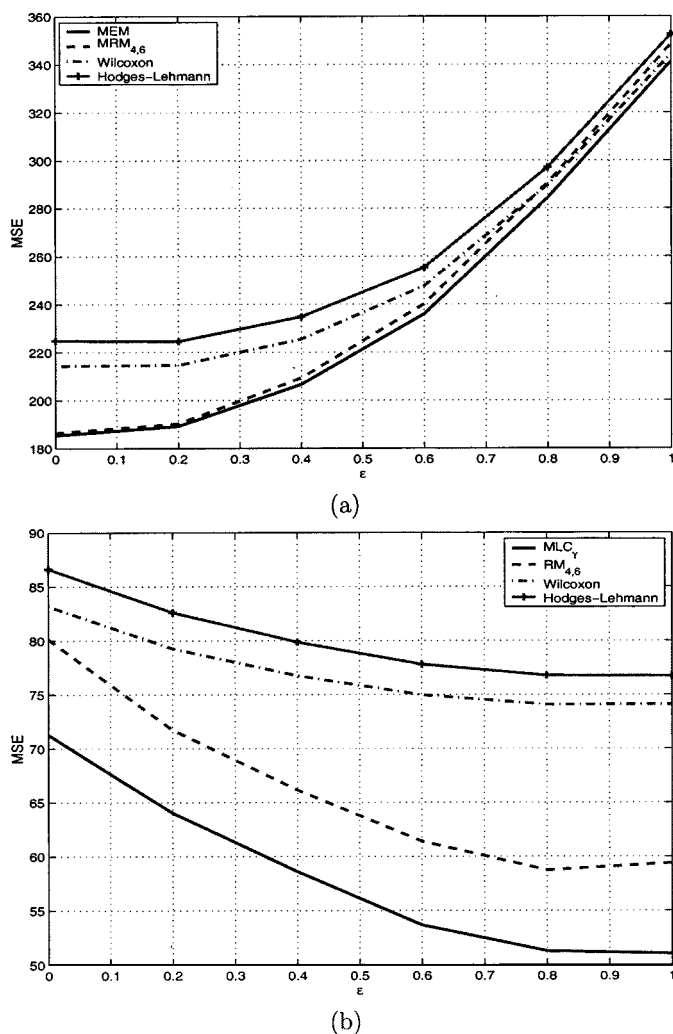


Fig. 6. Influence of the contamination fraction ϵ on filtering performance: MSE versus ϵ . (a) Gaussian and Laplacian noise. (b) Gaussian and α -stable noise.

by merely adjusting or optimizing the parameter λ . On the other hand, the filtering performance is also sensitive to the fraction of contamination ϵ . When $\epsilon = 0$, the mixed noise is purely Gaussian, and when $\epsilon = 1$, it is purely Laplacian. Fig. 6 shows the influence of the parameters ϵ on the filtering performance.

VII. CONCLUSION

Convex combinations of filtering methods using the mean, the relaxed median, and the LogCauchy filters were proposed. Some deterministic and statistical properties were studied, and the asymptotic analysis of the mean-median filter was performed. The decrease of the lower bound of the mean-relaxed median filter yields better detail preservation at a cost of a reduction in its noise reduction performance. It was shown that the proposed schemes are efficient in suppressing heavy-tailed as well as mixed noise, compared with other filtering techniques. In future work, we plan on extending the theoretical and experimental results to nonrecursive and weighted filtering

techniques, as well as on optimizing the performance of these filters by selecting their optimal parameters. The results will be reported elsewhere.

ACKNOWLEDGMENT

The authors would like to thank Dr. I. Schick for many fruitful comments regarding this work. The authors would also like to acknowledge the associate editor and the anonymous reviewers for helpful and constructive comments.

REFERENCES

- [1] J. Astola and P. Kuosmanen, *Fundamentals of Nonlinear Digital Filtering*. Boca Raton, FL: CRC, 1997.
- [2] P. D. Wendt, E. J. Coyle, and N. C. Gallagher, "Stack filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 898–919, Aug. 1986.
- [3] G. R. Arce and R. E. Foster, "Detail preserving ranked-order based filters for image processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 83–98, Jan. 1989.
- [4] R. Yang, L. Yin, M. Gabbouj, J. Astola, and Y. Neuvo, "Optimal weighted median filters under structural constraints," *IEEE Trans. Signal Processing*, vol. 43, pp. 591–604, Mar. 1995.
- [5] R. C. Hardie and K. E. Barner, "Rank conditioned rank selection filters for signal restoration," *IEEE Trans. Image Processing*, vol. 3, pp. 192–206, Mar. 1994.
- [6] A. Ben Hamza, P. Luque, J. Martinez, and R. Roman, "Removing noise and preserving details with relaxed median filters," *J. Math. Imag. Vision*, vol. 11, no. 2, pp. 161–177, Oct. 1999.
- [7] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proc. IEEE*, vol. 81, pp. 986–1010, July 1993.
- [8] S. Ambike and D. Hatzinakos, "A new filter for highly impulsive α -stable noise," in *Proc. Int. Workshop Nonlinear Signal Image Process.*, Greece, 1995.
- [9] J. B. Bednar and T. L. Watt, "Alpha-trimmed means and their relationships to median filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 145–153, Feb. 1987.
- [10] H. Krim, D. Tucker, S. Mallat, and D. Donoho, "On denoising and best signal representation," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2225–2238, Nov. 1999.
- [11] H. Krim and I. C. Schick, "Minimax description length for signal denoising and optimized representation," *IEEE Trans. Inform. Theory*, vol. 45, pp. 898–908, Apr. 1999.
- [12] P. Huber, "Robust estimation of a location parameter," *Ann. Math. Stat.*, vol. 35, pp. 1753–1758, 1964.
- [13] —, *Robust Statistics*. New York: Wiley, 1981.
- [14] L. A. Jaeckel, "Robust estimates of location: symmetry and asymmetry contamination," *Ann. Math. Stat.*, vol. 42, no. 3, pp. 1020–1034, 1971.
- [15] J. R. Collins, "Robust estimation of a location parameter in the presence of asymmetry," *Ann. Math. Stat.*, no. 4, pp. 68–85, 1976.
- [16] A. C. Bovik, T. S. Huang, and D. C. Munson, "A generalization of median filtering using linear order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1342–1350, Dec. 1983.
- [17] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI7, pp. 165–177, Mar. 1985.
- [18] E. E. Kuruoglu, C. Molina, S. J. Gosdill, and W. J. Fitzgerald, "A new analytic representation of the α -stable density function," in *Proc. Amer. Stat. Soc.*, 1997.
- [19] F. Steiner, *The Most Frequent Value*. Budapest, Hungary: Akademiai Kiado, 1991.
- [20] J. G. Gonzalez and G. R. Arce, "Optimality of the myriad filter in practical impulsive-noise environments," *IEEE Trans. Signal Processing*, vol. 49, pp. 438–441, Feb. 2001.
- [21] T. Chen, K. K. Ma, and L. H. Chen, "Tri-state median filter for image denoising," *IEEE Trans. Image Processing*, vol. 8, pp. 1834–1838, Dec. 1999.

A. Ben Hamza received the doctorate degree in applied mathematics from the University of Granada, Grenada, Spain. He is currently pursuing the Ph.D. degree in electrical engineering at North Carolina State University, Raleigh (NC State).

From March 2000 to February 2001, he was a Research Associate with Department of Electrical and Computer Engineering, NC State. His research interests include nonlinear probabilistic and variational filtering, information-theoretic measures, and applied differential geometry and topology.

Hamid Krim (M'80-SM'98) received all his degrees in electrical engineering.

As a Member of Technical Staff at AT&T Bell Labs, Murray Hill, NJ, he has worked in the areas of telephony and digital communication systems/subsystems. Following an NSF postdoctoral fellowship at Foreign Centers of Excellence, LSS/University of Orsay, Paris, France, he became a Research Scientist at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, performing and supervising research. He is presently on the faculty in the ECE Department of Electrical and Computer Engineering, North Carolina State University, Raleigh. His research interests are in statistical signal and image analysis and mathematical modeling with a keen emphasis on applied problems.