NONLINEAR IMAGE FILTERING IN A MIXTURE OF GAUSSIAN AND HEAVY-TAILED NOISE

A. Ben Hamza *Hamid Krim*

Department of Electrical and Computer Engineering
North Carolina State University, Raleigh, NC 27695-7914, USA
E-mails: {`abhamza, ahk`}`@eos.ncsu.edu`

ABSTRACT

Inspired by robust estimation, nonlinear denoising methods combining the mean, the median, and the LogCauchy filters are proposed. Some statistical and asymptotic properties are studied, and comparisons with other nonlinear filtering schemes are performed. Experimental results showing a much improved performance of the proposed filters in the presence of Gaussian and heavy-tailed noise are analyzed and illustrated.

1. INTRODUCTION

A variety of models have been sources in modeling impulsive noise inluding 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 that other models [1]. 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 which 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* [1].

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 [2]. This led to a number of other proposed nonlinear schemes to attain a balance between the two. Among these proposed filters, figure Wilcoxon and Hodges-Lehmann filters [2].

Approaches to wavelet-based denoising have generally relied on the assumption on Gaussian noise, and are therefore sensitive to outliers, i.e., to noise distributions whose tails are heavier than the Gaussian distribution, such as Laplacian distribution. For independent ϵ-contaminated Gaussian distributions of the wavelet coefficients, Krim and Schick [4] derive a robust estimator of the wavelet coefficients based on minimax description length.

In the next section, we provide a brief review of Huber minimax approach, some basic sliding window filters and symmetric α-stable ($S\alpha S$) distributions. In Section 3, a nonlinear filtering structure called *Mean-Median* filter is introduced and its asymptotic analysis is performed. Section 4 is devoted to another class of nonlinear denoising techniques called *Mean-LogCauchy* filters. Finally, in Section 5, we provide experimental results to show a much improved performance of the proposed filters at removing noise from images corrupted by ϵ-contaminated Gaussian and heavy tailed noise, while preserving well image structures.

This work was supported by an AFOSR grant F49620-98-1-0190 and by ONR-MURI grant JHU-72798-S2 and by NCSU School of Engineering.

2. BACKGROUND

Consider the additive noise model

$$X_{\boldsymbol{i}} = S_{\boldsymbol{i}} + V_{\boldsymbol{i}}, \qquad \boldsymbol{i} \in \mathbb{Z}^m, \tag{1}$$

where $\{S_{\boldsymbol{i}}\}$ be a discrete m-dimensional deterministic sequence corrupted by the zero-mean noise sequence $\{V_{\boldsymbol{i}}\}$, and $\{X_{\boldsymbol{i}}\}$ is the observed sequence. The objective is to estimate the sequence $S_{\boldsymbol{i}}$ based on a filtering output $Y_{\boldsymbol{i}} = \mathcal{F}(X_{\boldsymbol{i}})$, 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 of ϵ-contaminated normal neighborhood proposed by Huber [3]

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

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”. The presence of outliers in a nominally normal sample can be modeled here by a distribution H with tails heavier than normal. Note that symmetry ensures the unbiasedness of the maximum likelihood estimator, making the expression for its asymptotic variance considerably simpler. Krim and Schick [4] proposed a robust wavelet thresholding technique based on the minimax description length (MMDL) principle, determining the least favorable distribution in $\mathcal{P}_\epsilon$ 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 $W_{\boldsymbol{i}} = \{X_{\boldsymbol{i}+\boldsymbol{r}} : \boldsymbol{r} \in W\}$ to be the window centered at location $\boldsymbol{i}$. The output of the mean filter is given by

$$Y_{\boldsymbol{i}} = \overline{W_{\boldsymbol{i}}} = \operatorname*{argmin}_\theta \sum_{\boldsymbol{r}\in W} (X_{\boldsymbol{i}+\boldsymbol{r}} - \theta)^2. \tag{2}$$

where $\overline{W_{\boldsymbol{i}}}$ is the sample mean of the window $W_{\boldsymbol{i}}$.
Denote by $[W_{\boldsymbol{i}}]_{(k)}$ the k-th order statistic of the samples in $W_{\boldsymbol{i}}$, that is $[W_{\boldsymbol{i}}]_{(1)} \leq [W_{\boldsymbol{i}}]_{(2)} \leq \cdots \leq [W_{\boldsymbol{i}}]_{(2N+1)}$.
The output of the standard median (SM) filter is given by

$$Y_{\boldsymbol{i}} = [W_{\boldsymbol{i}}]_{(N+1)} = \operatorname*{argmin}_\theta \sum_{\boldsymbol{r}\in W} |X_{\boldsymbol{i}+\boldsymbol{r}} - \theta|. \tag{3}$$

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 [1]. 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 [1].
When $\alpha \in (0, 2)$, an $S\alpha S$ random variable has infinite variance, and the Cauchy ($\alpha = 1$) is the only distribution which 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 [5] is the maximum log-likelihood estimator of the location parameter for a Cauchy density, and yields the following

$$Y_{\boldsymbol{i}} = \mathrm{LC}_\gamma(W_{\boldsymbol{i}}) = \operatorname*{argmin}_{\theta} \sum_{\boldsymbol{r} \in W} \log\left(\gamma^2 + (X_{\boldsymbol{i}+\boldsymbol{r}} - \theta)^2\right), \quad (4)$$

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

3. THE MEAN-MEDIAN FILTER

From Eqs. (2) and (3), it can easily be seen that the mean filter is optimal for Gaussian noise in the sense of mean square error while 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 $\mathcal{P}_\epsilon$, i.e. $P = (1-\epsilon)G + \epsilon L$, 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 $L \in \mathcal{S}$). This assumption on the noise to be ϵ-contaminated Gaussian and Laplacian distributed is motivated by 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. 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*

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

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 and other reasons, 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 then given by [3]

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

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

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

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 median filters are given by [3]

$$IF(x; \overline{W_{\boldsymbol{i}}}, F_\theta) = x - \theta,$$

and

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

Then it follows that the influence function of the MEM filter is given by

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

Using (5) and (6), the following result holds.

Proposition 1 *The asymptotic variance $V(\mathrm{MEM}, F_\theta)$ of the MEM filter at the distribution F*

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

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

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

Minimizing (7) over λ, we obtain the minimum attainable asymptotic variance, and the filter attaining that minimum asymptotic variance will then provide the best filtering performance.

Corollary 1 *The minimum value of $V(\mathrm{MEM}, F_\theta)$ is attained at $\lambda_{\min}$ given by*

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

Example: If the input is i.i.d. $\mathcal{N}(\theta, \sigma^2)$, then using (8), we obtain $\lambda_{\min} \approx 2/(2+\pi)$.

4. MEAN-LOGCAUCHY FILTERS

The LogCauchy filter has been shown to outperform the standard median filter in removing highly α-stable noise [5], then the MEM filter can be improved replacing the median by the LogCauchy, and therefore a new class of nonlinear filters is derived.

Now we assume that the noise probability distribution P is a scaled version of a member of $\mathcal{P}_\epsilon$ 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 the characteristic function φ_ϵ of the random variable $(1-\epsilon)X_G + \epsilon X_S$ is given by

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

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 [6] 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 having 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 2 *The output of Mean-LogCauchy (MLC_γ) filter with parameter γ is given by*

$$Y_{\boldsymbol{i}} = \mathrm{MLC}_\gamma(W_{\boldsymbol{i}}) = (1-\lambda)\overline{W_{\boldsymbol{i}}} + \lambda\, \mathrm{LC}_\gamma(W_{\boldsymbol{i}}),\ \ \lambda \in [0,1], \quad (9)$$

where 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

$$\begin{aligned} \hat{\theta}_{\boldsymbol{i}} &= \operatorname*{argmax}_\theta \ell_\gamma(\theta; W_{\boldsymbol{i}}) \\ &= \operatorname*{argmax}_\theta \log \prod_{\boldsymbol{r}\in W} \frac{\gamma}{\pi}\left(\frac{1}{\gamma^2 + (X_{\boldsymbol{i}+\boldsymbol{r}} - \theta)^2}\right), \quad (10) \end{aligned}$$

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

It is clear that for a given γ, solving (10) is equivalent to minimizing the function $\rho_\gamma(\theta; W_{\boldsymbol{i}})$ given by

$$\rho_\gamma(\theta; W_{\boldsymbol{i}}) = \prod_{\boldsymbol{r}\in W}\left(\gamma^2 + (X_{\boldsymbol{i}+\boldsymbol{r}} - \theta)^2\right), \quad (11)$$

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 $\rho_\gamma(\theta; W_{\boldsymbol{i}})$. This is very important because $\rho_\gamma(\theta; W_{\boldsymbol{i}})$ is a polynomial of degree $2(2N+1)$ in θ and its characteristics can then be obtained easily. It can be shown that $\rho_\gamma(\theta; W_{\boldsymbol{i}})$ is a convex function of θ if $\gamma \geq [W_{\boldsymbol{i}}]_{(2N+1)} - [W_{\boldsymbol{i}}]_{(1)}$, and therefore has a unique minimum $\theta_0 \in [[W_{\boldsymbol{i}}]_{(1)}, [W_{\boldsymbol{i}}]_{(2N+1)}]$. At $\gamma = 0$, the function $\rho_\gamma(\theta; W_{\boldsymbol{i}})$ has distinct minima at all the points $X_{\boldsymbol{i}+\boldsymbol{r}}$. If γ is increased, the number of minima decreases. After a certain limit of γ, there is only a unique minimum.

Proposition 2 *When $\gamma \to \infty$, the Mean-LogCauchy filter becomes the mean filter, i.e.*

$$\mathrm{MLC}_\gamma(W_{\boldsymbol{i}}) \to \overline{W_{\boldsymbol{i}}} \quad \textit{as} \quad \gamma \to \infty.$$

Proof. Using basic properties of the argmin function, the output of the LogCauchy filter can be expressed as

$$\begin{aligned} \mathrm{LC}_\gamma(W_{\boldsymbol{i}}) &= \operatorname*{argmin}_\theta \sum_{\boldsymbol{r}\in W} \log\left(\gamma^2 + (X_{\boldsymbol{i}+\boldsymbol{r}} - \theta)^2\right) \\ &= \operatorname*{argmin}_\theta \sum_{\boldsymbol{r}\in W} \gamma^2 \log\left(1 + \frac{(X_{\boldsymbol{i}+\boldsymbol{r}} - \theta)^2}{\gamma^2}\right) \\ &= \operatorname*{argmin}_\theta \sum_{\boldsymbol{r}\in W} \log\left(1 + \frac{(X_{\boldsymbol{i}+\boldsymbol{r}} - \theta)^2}{\gamma^2}\right)^{\gamma^2} \end{aligned}$$

Since

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

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

$$\mathrm{LC}_\gamma(W_{\boldsymbol{i}}) \to \operatorname*{argmin}_\theta \sum_{\boldsymbol{r}\in W} (X_{\boldsymbol{i}+\boldsymbol{r}} - \theta)^2 \quad \text{as} \quad \gamma \to \infty.$$

This concludes the proof using (2) and (9). ■

Note that asymptotically, the tuning parameter γ transforms a nonlinear filter to a linear one.

5. EXPERIMENTAL RESULTS

This section presents simulation results where the proposed filters are applied to enhance images corrupted by mixed Gaussian and heavy tailed 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 vary well at the nominal model but badly at an outlier, while another do poorly at the 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 which is obviously tend to downplay the influence of large errors, compared to mean square error precisely in the presence of heavy-tailed noise.

Mean square error (MSE) between the filtered and the original image is evaluated to quantitatively compare the good performance of the proposed filters with other filtering techniques.

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. 1(a) was contaminated by both Gaussian white noise ($\sigma^2 = 100$) and α-stable noise $S\alpha S(\alpha = 0.5)$. The ϵ-contaminated mixed noise corrupted image is shown in Fig. 1(b). The visual comparison with other techniques is shown in Fig. 1. The relaxed median filter [7] outperforms Wilcoxon and Hodges-Lehmann in suppressing highly α-stable noise, while the Mean-LogCauchy filter, with mixture parameter $\lambda = \pi/(2+\pi)$ and optimal tuning parameter $\gamma = 2.38$, achieves the best performance. In the simulation results of Fig.1, the contamination fraction ϵ is chosen to be equal to λ.

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 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 α-stable. Fig. 2 shows the influence of the parameter ϵ on the filtering performance.

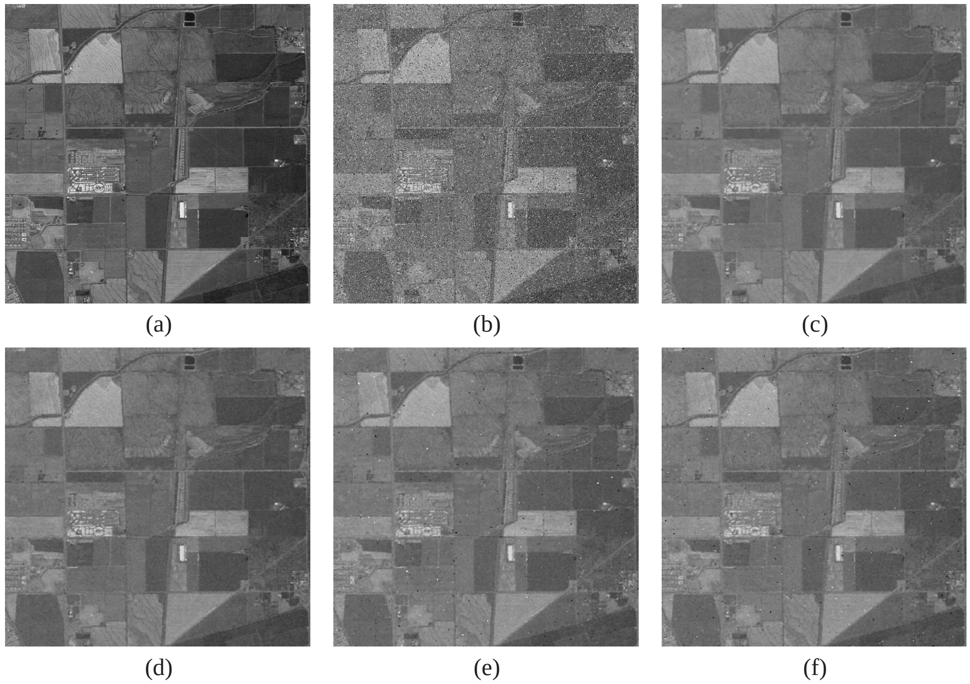

Fig. 1. Filtering results in the presence of ϵ-contaminated Gaussian and α-stable noise, and using a 3×3 square window: (a) Original image, (b) ϵ-mixed noisy image with $\mathcal{N}(0, 100)$ and $S\alpha S$, (c) Output of the MLC filter, $\lambda = 2/(2 + \pi)$, (d) Output of the relaxed median filter, (e) Output of the Wilcoxon filter, and (f) Output of the Hodges-Lehmann filter.

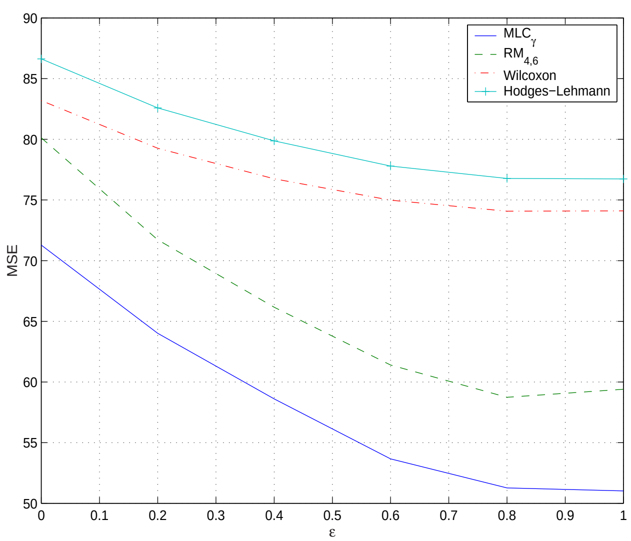

Fig. 2. Influence of the contamination fraction ϵ on filtering performance: MSE vs. ϵ.

6. REFERENCES

[1] M. Shao and C.L. Nikias, "Signal Processing with fractional lower order moments: Stable processes and their applications," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 986-1010, July 1993.

[2] J. Astola and P. Kuosmanen, *Fundamentals of Nonlinear Digital Filtering*, CRC Press LLC, 1997.

[3] P. Huber, *Robust Statistics*, John Wiley, 1981.

[4] H. Krim and I.C. Schick, "Minimax description length for signal denoising and optimized representation," *IEEE Trans. Information Theory*, vol 45, no. 3, pp. 898-908, April, 1999.

[5] S. Ambike and D. Hatzinakos, "A new filter for highly impulsive α-stable noise," *Proc. 1995 Int. Workshop Nonlinear Signal Image Processing*, Greece, 1995.

[6] E.E. Kuruoglu, C. Molina, S.J. Gosdill and W.J. Fitzgerald, "A new analytic representation of the α-stable density function," *American Statistical Society Proceedings*, 1997.

[7] A. Ben Hamza, P.Luque, J.Martinez, and R. Roman "Removing noise and preserving details with relaxed median filters," *Journal of Mathematical Imaging and Vision*, vol. 11, no. 2, pp. 161-177, October 1999.