SCALAR QUANTIZATION OF ALPHA-STABLE DISTRIBUTED RANDOM VARIABLES

Panagiotis Reveliotis, Panagiotis Tsakalides, and Chrysostomos L. Nikias

Signal & Image Processing Institute
Department of Electrical Engineering – Systems
University of Southern California
Los Angeles, CA 90089–2564
e-mail: tsakalid@sipi.usc.edu

ABSTRACT

Efficient stochastic data processing preassumes proper modeling of the statistics of the data source. This paper addresses the issues that arise when the data to be processed exhibits statistical properties which depart significantly from those implied under the Gaussianity assumption. This type of data has been found to be encountered in image, speech and other compression applications. For the cases under consideration, techniques based on the statistical theory of alpha-stable distributions have been found to give the most proper solution to the modeling problem. Furthermore, an alternative to the common mean-square error (MSE) quantizer for the efficient, by means of distortion minimization, scalar quantization of heavy-tailed data is presented. The proposed quantizer is based on a particular member of the family of alpha-stable distributions, namely the Cauchy distribution. The results of the performance of this quantizer when applied to simulated as well as real data are also presented.

1. INTRODUCTION

The problem of processing data in digital form is intrinsically convolved with the problem of quantization. The precise formulation of this problem was first addressed in the literature by Max in [1] for the case of the MSE criterion. Equivalent results are reported in [2]. These reports include the necessary conditions for the design of the optimal quantizer, by means of the minimization of the MSE. They also present the thresholds for the optimal quantizer in the case of data following the Gaussian distribution. The sufficient conditions for the optimal MSE quantizer were first investigated in [3] and further examined for a broader class of optimization criteria in [4]. In general, two approaches have been so far proposed in the literature for the solution of the optimal quantization problem. The first is based on an itiretative method for the solution of a system of equations, determining the stationary points of the distortion measure, which results in a locally optimum quantizer [1, 2]. The second method is based on a computationally heavily demanding search, using dynamic programming [5]. It is also worth noting that non symmetric quantization schemes have been found to be optimal for certain symmetric distributions and for symmetric error weighting functions [5, 6].

The work in this paper was supported by the Office of Naval Research under Contract N00014-92-J-1034 and by the Advanced Research Project Agency under Contract DABT63-95-C-0092.

Although the optimal quantization problem has been completely solved for data following certain distributions such as Gaussian, Laplacian or Rayleigh, there exist cases when the data does not follow any of these distributions, following instead a so-called heavy-tailed marginal distribution. This type of data have been observed in such diverse fields as telecommunications, finance and economics, radar and sonar, and speech and image compression. Examples include file lengths, cpu time to complete a job, inter-arrival times between packets in network communications [7, 8], stock returns and interest rate movements in economics [9], clutter returns in radar [10, 11], and coefficients in state-of-the-art image coders based on wavelets.

The problem of the optimal quantization of heavy-tailed data is still open and is being addressed in this paper. We demonstrate that alpha-stable distributions are sufficiently flexible and rich to appropriately model wavelet coefficients in image coding applications. Our modeling results give rise to new and challenging problems in information theory in general and quantization theory in particular and open new areas of mathematical research in rate distortion theory within the alpha-stable framework. In Section 2, we present some necessary preliminaries on alpha-stable processes and results on the modeling of wavelet coefficients by means of stable distributions. In Section 3.1, we formulate the quantization problem of heavy-tailed data and introduce the Cauchy Quantizer for sources following the Cauchy density. In Section 3.2, we compare the performance of our proposed scheme with the performance of the Gaussian and Laplacian quantizers by means of simulated and real data. Finally, in Section 4, we discuss some of the many issues that need to be addressed concerning the information theoretic aspects of the alpha-stable family.

2. DATA MODELING WITH ALPHA-STABLE DISTRIBUTIONS

In this section, we introduce the statistical model that will be used to describe sources of a heavy-tailed nature. The model is based on the class of symmetric α-stable ($S\alpha S$) distributions and is well-suited for characterizing distributions which exhibit heavy tails. A review of the state of the art on stable processes from a statistical point of view is provided by a collection of papers edited by Cambanis, Samorodnitsky and Taqqu [12], while textbooks in the area were written by Samorodnitsky and Taqqu [13], and by Nikias and Shao [14].

The appeal of $S\alpha S$ distributions as a statistical model for signals derives from some important properties. Namely, stable processes satisfy the stability property which states that linear combinations of jointly stable variables are indeed stable. They arise as limiting processes of sums of independent, identically-distributed random variables via the generalized central limit theorem. They are described by their characteristic exponent α, taking values $0 < \alpha \leq 2$. Gaussian processes are stable processes with $\alpha = 2$ while Cauchy processes result when $\alpha = 1$. In fact, no closed-form expressions for the general $S\alpha S$ probability density function (pdf) are known except for the Gaussian and the Cauchy members. Stable distributions have heavier tails than the normal distribution, possess finite pth order moments only for $p < \alpha$, and are appropriate for modeling signals with outliers.

The symmetric α-stable ($S\alpha S$) distribution is best defined by its characteristic function

$$\varphi(\omega) = \exp(\jmath\delta\omega - \gamma|\omega|^{\alpha}), \tag{1}$$

where α is the *characteristic exponent* restricted to the values $0 < \alpha \leq 2$, δ ($-\infty < \delta < \infty$) is the *location parameter*, and γ ($\gamma > 0$) is the *dispersion* of the distribution. For values of α in the interval $(1, 2]$, the location parameter δ corresponds to the mean of the $S\alpha S$ distribution, while for $0 < \alpha \leq 1$, δ corresponds to its median. The dispersion parameter γ determines the spread of the distribution around its location parameter δ, similar to the variance of the Gaussian distribution. The characteristic exponent α is the most important parameter of the $S\alpha S$ distribution and it determines the shape of the distribution.

Although the $S\alpha S$ density behaves approximately like a Gaussian density near the origin, its tails decay at a lower rate than the Gaussian density tails. While the Gaussian density has exponential tails, the stable densities have algebraic tails. *The smaller the characteristic exponent α is, the heavier the tails of the $S\alpha S$ density.* This implies that random variables following $S\alpha S$ distributions with small characteristic exponents are *highly impulsive.*

Figures 1 and 2 show results on the modeling of the statistics of wavelet coefficients by means of the Gaussian, Laplacian, and $S\alpha S$ distributions. The heavy-tailed nature of the data is obvious in Figure 1 which shows the wavelet coefficients for a given subband of the Lena image. The estimation of the parameters of the stable distribution from the coefficients was achieved by methods based on fractional lower-order moments, as described in [15]. For the particular case shown here, the characteristic exponent of the $S\alpha S$ distribution which best fits the data was calculated to be $\alpha \approx 1.3$. Figure 2 demonstrates that the $S\alpha S$ distribution is superior to the Gaussian or Laplacian distribution for modeling the particular wavelet coefficient data under study.

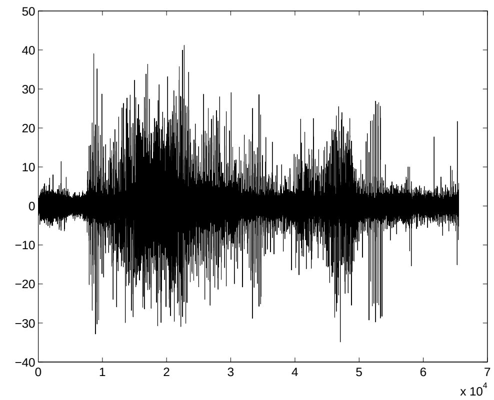

Figure 1: Wavelet coefficient series.

3. QUANTIZATION OF A CAUCHY SOURCE

3.1. Problem Formulation

In its general form, the problem of optimum scalar quantization can be considered as the task of defining the decision levels $d_0 < d_1 < \cdots < d_M$ and the reconstruction levels $r_1 < r_2 < \cdots < r_M$, in order to form the following partitioning of the data dynamic range $R = [L, U)$:

$$[L, U) = \bigcup_{k=0}^{M-1} [d_k, d_{k+1}), \tag{2}$$

and represent all the data values x lying within the subrange $[d_k, d_{k+1})$ with the reconstruction level r_k so that a distortion measure $D(e)$ is minimized, where e is the quantization error, defined by:

$$e = x - r_k. \tag{3}$$

In other words, e is the difference of the reconstruction level from the data value, which it represents. For stochastic data, the distortion measure is defined as the expected value of an error weighting function:

$$D(e) = E[f(e)] = \int_L^U f(e)\, p(x)\, \mathrm{d}x, \tag{4}$$

where $p(x)$ is the pdf of the data distribution and $f(e)$ is the error weighting function. For the specific case under consideration, of data following a $S\alpha S$ distribution with $1 \leq \alpha < 2$, in order to define completely the quantization problem one has to determine $p(x)$ as well as $f(e)$.

Given that the pdf of a general non-Gaussian $S\alpha S$ distribution cannot be defined in closed form, except for $\alpha = 1$

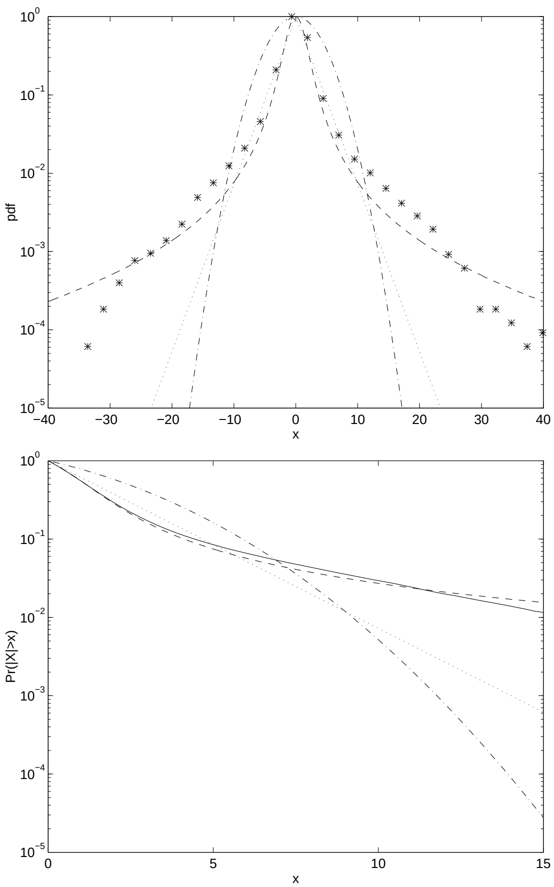

Figure 2: PDF comparison (top), stars: empirical, dash-dotted: Gaussian ($\sigma = 3.5780$), dashed: $S\alpha S$ ($\alpha = 1.3622$, $\gamma = 1.7101$), dotted: Laplacian ($\lambda = 0.4923$). Amplitude probability density (APD) comparison (bottom), solid: empirical.

the only available choice for $p(x)$ is the Cauchy probability density function:

$$p(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x-\mu)^2}, \tag{5}$$

where μ is the location parameter and γ is the dispersion. Note that if a Cauchy random variable (r.v.) X follows the distribution described by (5), then $\frac{X-\mu}{\gamma}$ is also a Cauchy r.v. with location parameter equal to zero and dispersion equal to one.

On the other hand, the choice of $f(e)$ is constrained by the fact that for Cauchy random variables, only moments of order less than one can be defined. Furthermore, $f(e)$ should be a symmetric and monotonically increasing function of e. For our analysis we have set

$$f(e) = \sqrt{|e|}, \tag{6}$$

which is a choice that satisfies the above mentioned conditions. An additional condition that the quantizer should be

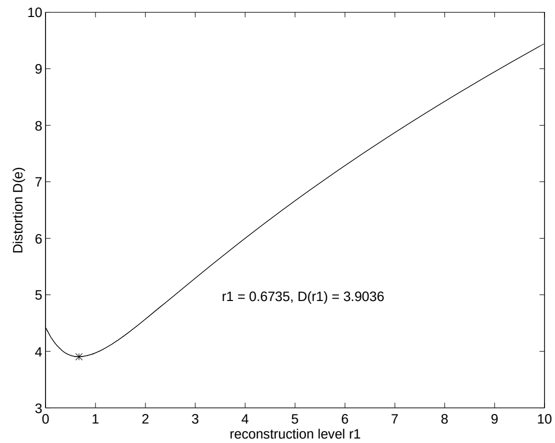

Figure 3: Plot of the distortion D as a function of the reconstruction level r_1 for $N = 2$ (relate with Table 1).

symmetric has been set, so that the mean of the quantization error for $S\alpha S$ random variables with $\alpha > 1$ is always zero. Note that for a Cauchy r.v. ($\alpha = 1$) the mean is not defined. For the symmetric quantizer due to the corresponding symmetry of $p(x)$ and $f(e)$ one decision level has to be set at zero and moreover the problem can be reduced to defining the quantizer for only positive values of data. The complete quantizer, having $N = 2M$ reconstruction levels, can be obtained by mirroring the defined thresholds (d_k and r_k) with respect to the y-axis.

Taking the above considerations into account, the quantization problem for $S\alpha S$ random variables can be formulated as follows: For a given number of levels M, determine the decision levels $d_0 < d_1 < \cdots < d_M$ and the reconstruction levels $r_1 < r_2 < \cdots < r_M$ so that

$$D = \sum_{k=1}^{M} \int_{d_{k-1}}^{d_k} \sqrt{|x - r_k|} \, \frac{1}{\pi} \frac{1}{1+x^2} \, \mathrm{d}x \tag{7}$$

where $d_0 = 0,\ d_M = \infty$, is minimized.

The quantization problem, as defined in (7), is a highly nonlinear optimization problem. The stationary points of the cost function $D(e)$ are given by the conditions:

$$\frac{\partial D}{\partial d_k} = 0,\ k = 1, \cdots, M-1 \quad \frac{\partial D}{\partial r_k} = 0,\ k = 1, \cdots, M. \tag{8}$$

It can be easily seen, that these conditions result respectively in the following relationships for d_k and r_k:

$$d_k = \frac{r_k + r_{k+1}}{2},\ k = 1, \cdots, M-1 \tag{9}$$

$$\int_{d_{k-1}}^{r_k} \frac{1}{\sqrt{r_k - x}} \frac{1}{1+x^2} \, \mathrm{d}x - \int_{r_k}^{d_k} \frac{1}{\sqrt{x - r_k}} \frac{1}{1+x^2} \, \mathrm{d}x = 0, \tag{10}$$

for $k = 1, \cdots, M$.

Equation (10) is a nonlinear integral equation for r_k. Hence, numerical iterative methods have been applied for

k	$N=2$ d_k	$N=2$ r_k	$N=4$ d_k	$N=4$ r_k
1	∞	0.6735	1.4270	0.4719
2			∞	2.3821
D	3.90368		3.30769	
H	1.000		1.964	

k	$N=8$ d_k	$N=8$ r_k	$N=16$ d_k	$N=16$ r_k
1	0.7036	0.3058	0.3806	0.1818
2	1.9588	1.1015	0.8380	0.5794
3	6.0641	2.8161	1.4913	1.0966
4	∞	9.3122	2.5885	1.8860
5			4.7687	3.2909
6			10.1505	6.2465
7			29.7345	14.0544
8			∞	45.4145
D	2.67460		2.06521	
H	2.854		3.720	

Table 1: Parameters for the Quantizer.

the solution of the highly nonlinear system of equations (9-10). The resulting values of the thresholds (d_k and r_k), for various values of the total number of reconstruction levels $N = 2M$, are shown in Table 1. This table shows also the values of the distortion D, as well as the entropy H achieved by the presented quantization schemes. It should be noted that the values of both the decision and reconstruction levels of this quantizer are significantly greater in absolute value than the corresponding levels of the Lloyd-Max quantizer for the Gaussian distribution, as they account for the much heavier tails of the Cauchy distribution.

The values of thresholds in Table 1 have been found to provide a locally optimal quantizer. This is indicated in Figure 3 as well as in Figure 4. When $N = 4$, the distortion D is a function of three variables, namely the two reconstruction levels r_1 and r_2 as well as the decision level d_1. Setting $d_1 = (r_1 + r_2)/2$ yields according to (9) all the candidate triplets (r_1, r_2, d_1) for being the optimal points. By these means distortion D can be considered as a function of just the two variables r_1 and r_2. Figure 4 depicts the contour plot of this function $D(r_1, r_2)$ indicating the local minimum for the values of r_1 and r_2 given at Table 1. It must be pointed out that further theoretical investigation is needed in order to determine whether the presented quantization schemes are also absolutely optimal.

3.2. Experimental Results

In the experimental part, the performance of the above presented quantizer was tested in comparison with the performance of both the optimal Mean Square Error (MSE) Gaussian and Laplacian quantizers, when applied on the same data.

Given that the Cauchy-based quantizer (hereto denoted as Cauchy Quantizer for simplicity) has been constructed

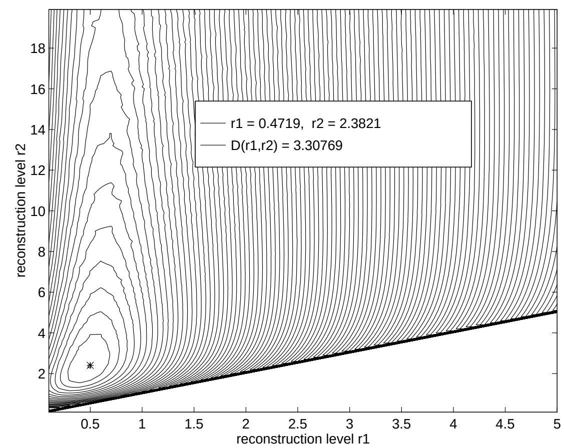

Figure 4: Contour plot of the distortion D as a function of the reconstruction levels r_1 and r_2 for $N = 4$ (relate with Table 1).

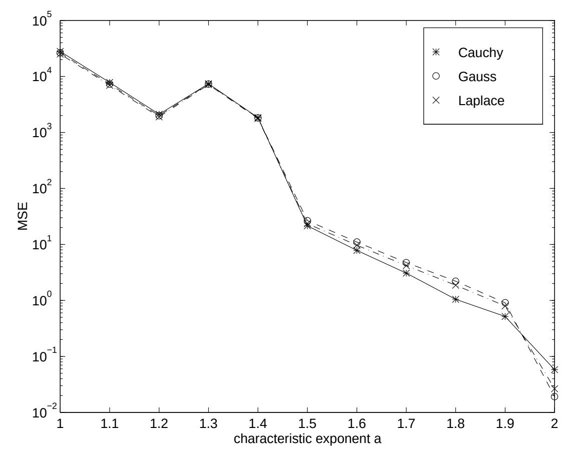

Figure 5: MSE for simulated data.

with reference to data following $S\alpha S$ distributions, at a first stage, simulated data were generated for $\alpha = 1, 1.1, \ldots, 2$. In each case the statistical parameters of the generated data, namely the mean and the standard deviation, as well as the location parameter and the dispersion, were estimated and the thresholds of the three quantizers for the standard distributions were scaled and translated in order to fit the data distribution. The data was then quantized according to each of the three quantization schemes, using quantizers with $N = 16$ levels. Based on the quantized data and the original data the following measures of distortion were computed:

- Mean Square Error (MSE), defined as:

$$MSE = \frac{1}{L} \sum_{i=1}^{L} (x_i - \hat{x}_i)^2, \quad (11)$$

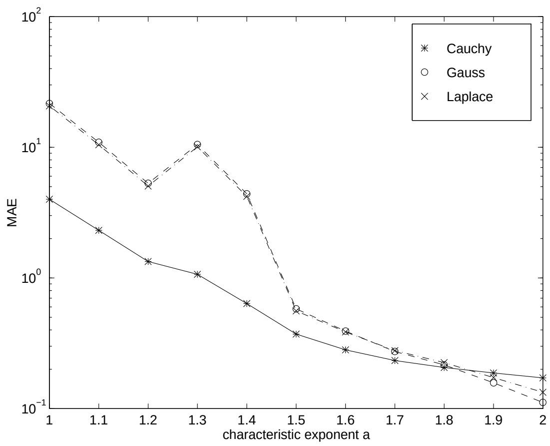

Figure 6: MAE for simulated data.

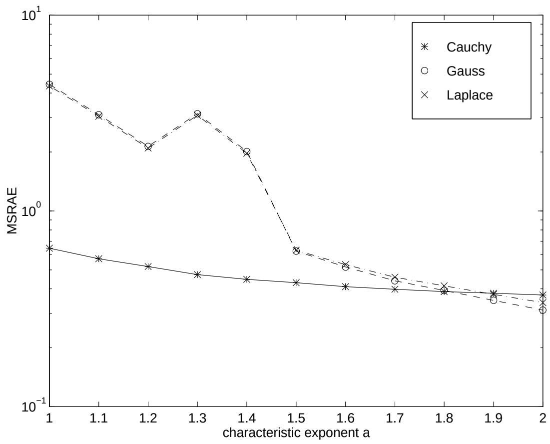

Figure 7: $MSRAE$ for simulated data.

- Mean Absolute Error (MAE), defined as:

$$MAE = \frac{1}{L} \sum_{i=1}^{L} |x_i - \hat{x}_i|, \tag{12}$$

- Mean Square Root Absolute Error (MSRAE), defined as:

$$MSRAE = \frac{1}{L} \sum_{i=1}^{L} \sqrt{|x_i - \hat{x}_i|}, \tag{13}$$

where $\hat{x}_i$ is the quantized value of the data value x_i and L is the number of generated data. The results for each quantization scheme, for each distortion measure, and for each value of α for $L = 65000$ are shown in Figures 5, 6, and 7.

At a second stage, instead of simulated, the wavelet coefficient data shown in Figure 1 were quantized using the Cauchy, Gaussian and Laplacian quantizers with $N = 2, 4, 8, 16$ reconstruction levels. Once again the above mentioned three distortion measures were computed in each case. The results are shown in Figures 8, 9, and 10.

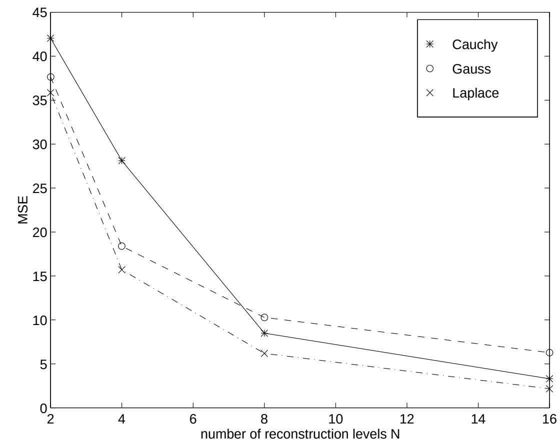

Figure 8: MSE for real data.

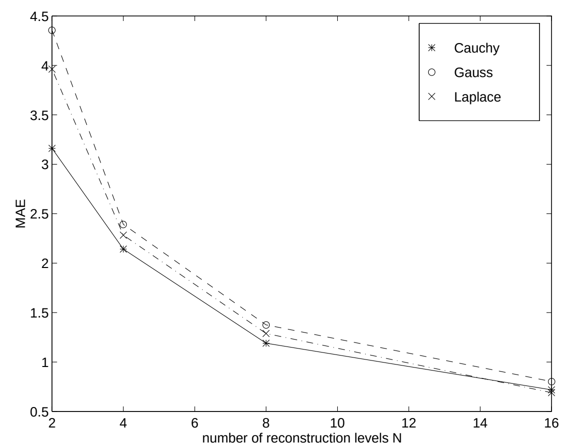

Figure 9: MAE for real data.

Figure 5 is mainly determined by the fact that $S\alpha S$ r.v.'s with $\alpha < 2$ have in theory infinite variance. This results in the great order of magnitude for MSE, especially for $\alpha \leq 1.5$ for the three quantization schemes. This figure implies that MSE is not a suitable measure of distortion for $S\alpha S$ r.v.'s with $\alpha < 2$. However, note that the optimality of the Gaussian MSE quantizer is evident for $\alpha = 2$.

Figure 6 on the other hand shows that the Cauchy quantizer achieves better performance, with respect to MAE, for $S\alpha S$ distributions which significantly depart from the Gaussian case (i.e. for values of $\alpha \leq 1.8$). Note that MAE is well defined for $\alpha > 1$, since $S\alpha S$ r.v.'s with $\alpha > 1$ have finite first moments. Moreover, MAE seems to be a more objective measure of distortion than MSE and $MSRAE$, since it weights all errors with the same factor. On the contrary MSE overestimates errors with absolute value greater than one and underestimates errors with absolute value less than one, while $MSRAE$ performs in exactly the opposite way.

Figure 7 shows the better performance of the Cauchy quantizer with respect to $MSRAE$ for the same range of

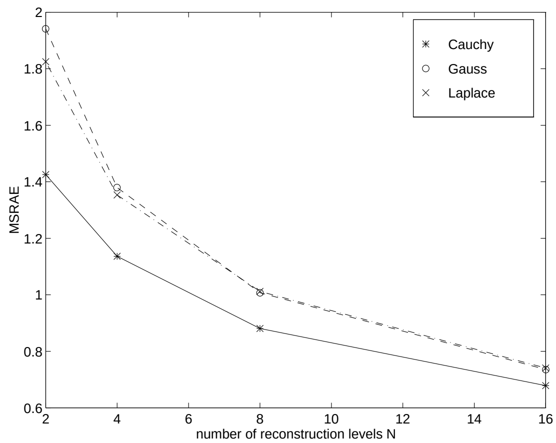

Figure 10: $MSRAE$ for real data.

α as in the MAE case. One should note that for $\alpha \geq 1.9$ Gaussian MSE quantizer achieves the best performance, which implies that optimality is determined by the distribution and not by the error weighting function, that was used in the design of the quantizer.

Coming to the real data, Figure 8 implies that our data lack the presence of very extreme outliers, resulting in more modest values for the MSE. However, the non Gaussian nature of the data is evident in the failure of the Gaussian quantizer to achieve a good performance, especially for $N = 8$ and $N = 16$. In Figure 9, it is shown that despite the lack of very extreme outliers, the Cauchy quantizer achieves a very good performance, with respect to MAE, for every number of reconstruction levels. Figure 10 depicts the $MSRAE$ and the superiority of the Cauchy quantizer is one more indication of the non Gaussian nature of the data.

4. DISCUSSION

As shown through the experiments with simulated as well as real data, the so-called Cauchy Quantizer appears to be a useful tool for discretizing data which follow distributions close to $S\alpha S$ with α significantly less than two.

There are certainly some further issues that should be addressed in future work. As already mentioned, theoretical issues concerning the uniqueness of the solutions of the highly nonlinear system of equations (9-10) and the absolute optimality of the quantizer are still open. Furthermore, from a practical point of view, the performance of the three examined quantization schemes should be also subjectively evaluated through the reconstruction of the images, after the quantization of their wavelet coefficients.

Another important issue to be examined concerns the effects of the use of other error weighting functions for the Cauchy quantizer, having the general form $f(e) = |e|^p$, $0 < p < 1$ (we have examined the case for $p = \frac{1}{2}$), especially for values of p close to one. Finally, the comparison of the Cauchy quantizer with the MAE-based (instead of MSE-based) Gaussian and Laplacian quantizers would also be useful, in order to examine the relative importance of the error weighting function and the probability density function in determining the values of the thresholds.

5. REFERENCES

[1] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, vol. 6, pp. 7–12, Mar. 1960.

[2] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. 28, pp. 129–137, Mar. 1982.

[3] P. E. Fleischer, "Sufficient conditions for achieving minimum distortion in a quantizer," in *IEEE Int. Conv. Rec.*, pp. 104–111, 1964.

[4] A. V. Trushkin, "Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions," *IEEE Trans. Inform. Theory*, vol. 28, pp. 187–198, Mar. 1982.

[5] D. K. Sharma, "Design of absolutely optimal quantizers for a wide class of distortion measures," *IEEE Trans. Inform. Theory*, vol. 24, pp. 693–702, Nov. 1978.

[6] P. Kabal, "Quantizers for the gamma distribution and other symmetrical distributions," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 32, pp. 836–841, Aug. 1984.

[7] S. Resnick, "Heavy tail modeling and teletraffic data," *Preprint*, 1995.

[8] W. Willinger, M. S. Taqqu, W. E. Leland, and D. V. Wilson, "Self-similarity in high-speed packet traffic: Analysis and modeling of ethernet traffic measurements," *Stat. Science*, vol. 10, pp. 67–85, 1995.

[9] P. Bidarkota and J. H. McCulloch, "Real stock returns: Non-normality, seasonality, and volatility persistence, but no predictanility," *Working Paper*, Sept. 1996.

[10] P. Tsakalides, R. Raspanti, and C. L. Nikias, "Joint target angle and doppler estimation in interference modeled as a stable process," in *Proc. 30th Conference on Information Sciences and Systems*, (Princeton, NJ), March 20-22 1996.

[11] P. Tsakalides, R. Raspanti, and C. L. Nikias, "Joint target angle and doppler estimation in stable impulsive interference," *IEEE Trans. Aerosp. Electron. Syst.* Submitted for publication consideration.

[12] S. Cambanis, G. Samorodnitsky, and M. S. Taqqu, eds., *Stable Processes and Related Topics*. Boston: Birkhauser, 1991.

[13] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. New York: Chapman and Hall, 1994.

[14] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*. New York: John Wiley and Sons, 1995.

[15] X. Ma and C. L. Nikias, "Parameter estimation and blind channel identification for impulsive signal environments," *IEEE Trans. Signal Processing*, vol. 43, pp. 2884–2897, December 1995.