Estimation of parameters for generalized Gaussian distribution

Conference Paper · May 2014

DOI: 10.1109/ISCCSP.2014.6877892

4 authors:

Alexey Roenko
KhAI - Aerospace university
38 PUBLICATIONS 128 CITATIONS

Vladimir Vasilyevich Lukin
KhAI - Aerospace university
637 PUBLICATIONS 5,434 CITATIONS

Igor Djurovic
University of Montenegro
251 PUBLICATIONS 3,291 CITATIONS

Marko Simeunović
University of Donja Gorica
57 PUBLICATIONS 349 CITATIONS

ESTIMATION OF PARAMETERS FOR GENERALIZED GAUSSIAN DISTRIBUTION

A.A. Roenko[1], V.V. Lukin[1], I. Djurović[2], M. Simeunović[2]

[1]National aerospace university named after N.E. Zhukovskiy «KhAI», Kharkov, Ukraine
e-mails: roenkomail@mail.ru, lukin@ai.kharkov.com
[2]University of Montenegro, Podgorica, Montenegro
e-mail: igordj@ac.me, markos@ac.me

ABSTRACT

Shape parameter estimation procedures for generalized Gaussian distribution are considered. It is shown that the existing estimators can be divided into four groups: maximum likelihood algorithm; moment-based methods; entropy matching estimators and global convergence algorithm. Besides, properties of two recently introduced estimators of shape parameter are discussed. They are based on the combination of two procedures that use the evaluation of the fourth central moment and robust measure of kurtosis. Statistical properties of all considered estimators are investigated by means of defining their bias and variance values for samples of sizes 1000 and 4000 elements and shape parameter values ranging from 0.3 to 2.

Index Terms— generalized Gaussian distribution, shape parameter estimation

1. INTRODUCTION

Effectiveness improvement of many modern systems and signal processing methods is directly connected with possibility to accurately model probability density function (pdf) of processed data. However, this distribution is often unknown. Such a situation is typical, e.g., in image compression and filtering, automatic methods for noise variance estimation, etc. [1]. Note that nowadays such methods are mostly based on different orthogonal transforms [2]. As a result, accurate description of distribution of spectral coefficients becomes the most important task [1, 3, 4].

During rather long period of time, a conventional solution was to apply Gaussian or Laplace pdf. However, at the end of 90th of the XX century it was shown that so-called generalized Gaussian distribution (GGD) could be used to accurately describe behavior of DWT and DCT coefficients for many test images in opposite to the Laplace pdf [5, 6]. In addition to earlier mentioned tasks, nowadays GGD is widely used in video processing [2], for embedding watermarks in images [7], for modeling speech signals [8]. GGD includes both bell-shaped distributions and pdfs with peaky shape of its maximum; it is symmetrical with respect to the location parameter. The form of pdf can be uniquely described by two parameters: shape parameter p and scale parameter σ [9]. Thus, the GGD pdf tail heaviness can vary in wide limits. This property also demonstrates importance and necessity of evaluating the shape parameter in order to effectively describe properties of data to be processed.

This paper deals with accuracy comparison for two recently introduced estimators of GGD shape parameter with the well-known p estimators. Performance analysis is done by calculating estimates' bias and variance for sample sizes corresponding to that of the DCT-based methods.

2. GENERALIZED GAUSSIAN MODEL

Analytically GGD is written as [9]:

$$f(x) = \frac{p \cdot A(p,\sigma)}{2\Gamma(1/p)} \exp\left(-A(p,\sigma)|x-\mu|\right)^{p} \qquad (1)$$

where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ denotes gamma-function; $A(p,\sigma) = \sigma^{-1}\sqrt{\Gamma(3/p)/\Gamma(1/p)}$, μ is a location parameter (typically equals to zero), σ and p denote scale and shape parameters, respectively.

Possible p values belong to the range (0; +∞). There are several particular cases: p=2 corresponds to the Gaussian distribution; GGD with p=1 is the Laplace pdf; distribution with p approaching +∞ asymptotically tends to the uniform one [9]. If $p \le 1$, GGD has peaky form of its maximum (i.e. there is a derivative discontinuity at this point) whereas for $p>1$ the pdf maximum is bell-shaped. In general, the smaller the p value is (for $p \le 2$), the heavier tails pdf has.

3. SHAPE PARAMETER ESTIMATION METHODS

3.1. Proposed estimators

Recently, we introduced four new p estimators. Two of them are based on the fourth central moment ($p_{\mu 4}$) and the robust measure of kurtosis (the so-called percentile coefficient of kurtosis (p_{PCK})). The estimators were obtained due to monotonically increasing or decreasing dependences of the mentioned parameters on parameter p for stochastic process with GGD [10].

Consider a data sample $\{x_1, x_2, \ldots, x_N\}$ of observed signal X which is generalized Gaussian distributed with parameters μ=0 and σ=1. Then, $p_{\mu 4}$ and p_{PCK} can be calculated as follows:

$$\mu_4 = \frac{1}{N}\sum_{n=1}^{N}\left(x_n - m_x\right)^4 = D(p),\ p_{\mu 4} = D^{-1}\left(\hat{\mu}_4\right), \quad (2)$$

$$\mu_{PCK} = \frac{1}{2}\frac{Q_3 - Q_1}{P_{90} - P_{10}} = C(p),\ p_{PCK} = C^{-1}(\hat{\mu}_{PCK}) \quad (3)$$

where Q_1 and Q_3 are the first and third quartiles, P_{90} and P_{10} denote the 90th and 10th percentiles, respectively. Note that the forms of $D(p)$ and $C(p)$ as well as detailed analysis of the estimators (2) and (3) can be found in [10].

Further investigations for increasing the accuracy of parameter p estimation have led us to an idea to combine estimators (2) and (3) in order to use positive features of each of them. As a result, two new estimators were obtained. The first one is based on "hard thresholding" [10]:

$$p_{TH} = \begin{cases} p_{PCK}, & \text{if } p_{\mu 4} \geq 1 \\ p_{\mu 4}, & \text{if } p_{\mu 4} < 1 \end{cases}. \quad (4)$$

The second proposed estimator appears as a weighted combination of $p_{\mu 4}$ and p_{PCK} values [10]:

$$p_{CMB} = \begin{cases} p_{\mu 4}, & \text{if } p_{\mu 4} > 2 \\ \left(\frac{2 - p_{\mu 4}}{2}\right) p_{PCK} + \frac{p_{\mu 4}}{2} p_{\mu 4}, & \text{if } p_{\mu 4} \leq 2 \end{cases}. \quad (5)$$

The main idea for both estimators is the same: when p<1, the estimator $p_{\mu 4}$ possesses smaller variance with respect to p_{PCK}. When $p \geq 1$, the situation is the opposite and better estimation accuracy can be achieved by applying p_{PCK} (see [10] for more details).

There also exist many other procedures for estimating GGD parameters. Most shape parameter estimators can be classified into four groups. They are: a) maximum likelihood estimator; b) moment-based methods; c) global convergence method; d) entropy matching estimators.

3.2. Maximum likelihood estimator (MLE)

For i.i.d. components of X, the likelihood function of the sample set is defined as [9]:

$$L(X;p,\sigma) = \log\prod_{n=1}^{N} f(x_n;p,\sigma) = \sum_{n=1}^{N}\log f(x_n;p,\sigma). \quad (6)$$

The maximum likelihood solution for the shape parameter can be obtained by taking the first derivative from (6) with respect to p and then setting it equal to 0 [9]:

$$g(p) = \partial L(X;p,\sigma)/\partial p = 0. \quad (7)$$

The $g(p)$ expression belongs to the class of transcendental equations and has no analytical solution. Many researchers propose to solve (7) using Newton-Raphson iteration method (see, for example, [8, 11]):

$$p_{k+1} = p_k - g(p_k)/g'(p_k),\ k = 0,1,2,... \quad (8)$$

where $g'(p_k)$ denotes the derivative of $g(p_k)$. The exact expressions for $g(p_k)$ and $g'(p_k)$ can be found in [12].

However, there are several drawbacks of such a solution. First, we have noticed that sometimes a rather great number of iterations (hundreds) are needed to achieve necessary precision due to slow convergence of (8). Second, calculation of $g'(p_k)$ and $g(p_k)$ is computationally demanding. Third, generally, the algorithm (8) is locally convergent, i.e. the initial p_0 value should be rather close to the true value (p_{true}) in order to $p_k \to p_{true}$ when $k \to \infty$. A standard approach for defining the initial value of Newton-Raphson algorithm is to apply the estimate from one of the moment-based methods (described below) [8, 11]. However, we noticed that in practice such estimates can be rather far from the true value of shape parameter. In these cases, the algorithm (8) can diverge and result in incorrect values.

In order to overcome these difficulties, we applied the Regula-Falsi iteration method based on the Secant algorithm [13]. Its main idea is in replacing the derivative $g'(p_k)$ by a finite difference approximation using a very small step size:

$$g'(p_k) = \left[g(p_k) - g(p_{k-1})\right]/\left[p_k - p_{k-1}\right]. \quad (9)$$

Then, for given p_0 and p_1, the algorithm is [13]:

$$p_{k+1} = p_k - \frac{p_k - p_{k-1}}{g(p_k) - g(p_{k-1})} g(p_k),\ k = 1,2,.... \quad (10)$$

The following condition should be satisfied for function $g(p)$ values [13]: $g(p_{k-1})g(p_k)<0$. One of the question of the Regula-Falsi iteration method is how to choose initial algorithm values p_0 and p_1. In [14], it was proposed to estimate p_0 by one of the moment-based methods. Then p_1 is found from the set $\{p_0 \pm l\Delta_l\}$ (l=1, 2,…, Δ_l is a predefined step size typically equal to 0.25) with the condition $g(p_0)g(p_1)<0$. Usually p_1 is calculated within one or two steps.

3.3. Moment-based estimators

The r-th order absolute moment for a signal with GGD can be evaluated as [6, 9]:

$$m_r = \frac{1}{N}\sum_{n=1}^{N}\left|x_n\right|^r. \quad (11)$$

whereas the GGD r-th absolute moment equals to [9]:

$$M_r = \int_{-\infty}^{\infty} |x|^r f(x,p,\sigma)dx = A^{-r}(p,\sigma)\frac{\Gamma((r+1)/p)}{\Gamma(1/p)}. \quad (12)$$

The main idea of the moment-based class of estimators is that the moments of sample set equal to the moments of pdf, i.e. $m_r=M_r$ [9]. Generally, any estimator from the considered class can be represented by two moments with the following generalized formula [6, 15]:

$$\frac{M_{r1}}{(M_{r2})^{r_1/r_2}} = \frac{\Gamma((r_1+1)/p)}{\Gamma^{r_1/r_2}((r_2+1)/p)\Gamma^{(1-r_1/r_2)}(1/p)} = \frac{m_{r_1}}{(m_{r_2})^{r_1/r_2}} \quad (13)$$

where r_1 and r_2 denote moment orders.

The estimator proposed in [6] (MRM) is based on the first and second absolute moments, i.e. r_1=1 and r_2=2:

$$R(p) = \frac{m_1}{\sqrt{m_2}} = \sqrt{\frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}},\ p_{MRM} = R^{-1}\left(\frac{\widehat{m}_1}{\sqrt{\widehat{m}_2}}\right). \quad (14)$$

Another moment-based estimator (KRM) proposed in [12, 15] and is obtained by setting r_1=2 and r_2=4.

3.4. Global convergence method (GCM)

Another estimation procedure is proposed in [16]. It is based on the assumption that if X obeys GGD, then

$$Z(p) \triangleq M_{2p} / (M_p)^2 - (1+p) = 0 \quad (15)$$

where $Z(p)$ is a convex function and (15) has unique root in (0; ∞). It is pointed out in [16] that the consistent estimator of $Z(p)$ can be given by

$$Z_N(p) \triangleq m_{2p} / m_p^2 - (1+p), \quad (16)$$

so the root of $Z_N(p)$=0 can be taken as the estimate of p. The Newton-Raphson iteration algorithm is proposed in [12, 16] for finding the root of $Z_N(p)$=0:

$$p_{k+1} = p_k - Z_N(p_k)/Z_N'(p_k),\ k = 0,1,2,\ldots. \quad (17)$$

However, this approach has the same difficulties which were described for the MLE case. That is why, we propose to apply the Regula-Falsi method for solving (15) as well.

3.5. Entropy matching estimators (EME)

Another class of estimators relies on matching the entropy of GGD distribution with that of a set of empirical data. Differential entropy of GGD signal is defined as [17]:

$$H(x) = \log_2\left(\frac{2}{p}\sqrt{\frac{\Gamma^3(1/p)}{\Gamma(3/p)}}\right) + \frac{1}{p\log 2}. \quad (18)$$

Suppose $H(X)$ is the entropy value of the data obtained at the output of an optimum entropy constrained uniform threshold quantizer with step size Δ [17]. Then $H(X)=H(x)-\log_2\Delta$ and substituting the expression for $H(x)$ (18) leads to

$$H(X) - \frac{1}{2}\log_2\frac{m_2}{\Delta^2} = \tilde{H}_2(p) = H(x) \quad (19)$$

where the right-hand side (i.e. $H(x)$) solely depends on the shape factor. Parameter p can be obtained as (EME2):

$$p_{EME2} = \tilde{H}_2^{-1}\left(\widehat{H}(X) - \frac{1}{2}\log_2\frac{\widehat{m}_2}{\Delta^2}\right) \quad (20)$$

where $\widehat{H}(X)$ and $\widehat{m}_2$ are the entropy and the second-order moment, respectively, calculated for the data sample. In [8] such an approach was generalized as:

$$\widehat{p}_{EME\,r} = \tilde{H}_r^{-1}\left(\widehat{H}(X) - \frac{1}{r}\log_2\frac{\widehat{m}_r}{\Delta^r}\right) \quad (21)$$

which includes (20) as a special case for r=2. It is shown in [8] that $\tilde{H}_2(p)$ has its maximal value at p=2 whereas $\tilde{H}_3(p)$ and $\tilde{H}_4(p)$ are the monotonically increasing functions at the interval of p values (0; 3). In this paper along with the case r=2, we will examine the estimators obtained for r=3 (EME3) and r=4 (EME4).

4. PERFORMANCE EVALUATION

Analysis is carried out for samples of sizes N=1000 and N=4000 elements and the total number of realizations in one experiment equals to 1000. These values correspond to the situation of processing an image 512x512 pixels by one of the DCT-methods in non-overlapping 8x8 blocks [10].

At first, let us consider the accuracy of two proposed estimators TH and CMB. Fig. 1 depicts the variance values of p_{TH} and p_{CMB} obtained for N=4000 as well as the characteristics of p_{MLE} as an accuracy limit that can be achieved by p estimation methods. The analysis shows that the CMB estimator provides better results compared to p_{TH} for all considered p values. Therefore, the TH estimator will not be considered further. Note that p_{CMB} loses MLE not more than 2 times for all p values except p=1.4 and 1.5.

Alongside with the variance values, we investigated estimators' bias. Let us examine this characteristic for EME (Fig. 2). Results demonstrate that all considered EM-procedures become biased: EME2 for all p>0.9, EME3 and EME4 for p>1.2. It can be explained by great steepness of the functions $\tilde{H}_r(p)$ for p>0.9. If the evaluated entropic function value (left part of Eq. (19)) even slightly deviates from the correct value, this will result in great bias value of p-estimator. It is also worth noting possible ambiguity in estimating p value in the case of EME2. Function $\tilde{H}_2(p)$ gets its maximal value for p=2 and then begins decreasing. It means that the same function value corresponds to two different p values at the interval [1.4; 3]. Note that the following thresholding was done in order to overcome

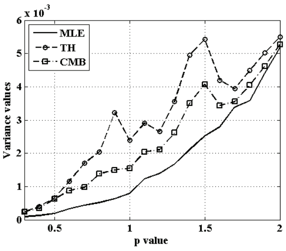

Fig. 1. Variance for MLE and estimators TH and CMB

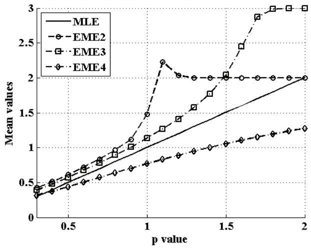

Fig. 2. Bias properties of the EME with r=2, 3 and 4

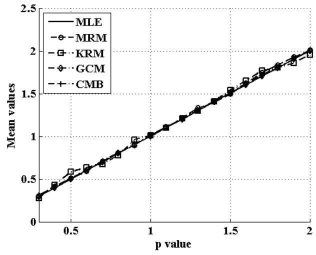

Fig. 3. Bias properties of considered estimators

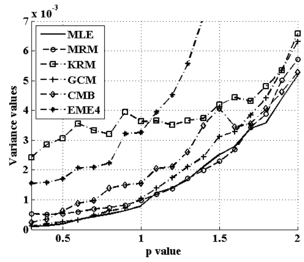

Fig. 4. Variance values of considered estimators

possible exception situations. If calculated function value was greater that its maximal value, than it was assigned to the maximal value. This explains the horizontal nature of EME2 and EME3 characteristics for p>1.2 and p>1.7, respectively. Based on the obtained results, only the EME4 estimator will be investigated further.

Results depicted in Fig. 3 show that other considered methods are practically unbiased. Accuracy analysis of the considered estimators (Fig. 4) shows that the best accuracy is achieved by MLE. Almost the same results for p≤1 are produced by GCM. For p>1, the GCM provides slightly worse accuracy but it loses not more than 20%. Remember that both MLE and GCM estimators require iterative procedures and are computationally demanding. Among the moment-based estimators, the better accuracy is achieved by MRM. Its variance is slightly greater than variance of MLE and GCM for p≤1 but almost coincides with the MLE accuracy for p>1. The KRM accuracy is the worst among the considered estimators almost for all p values. The only exception is the EME4 estimator which loses KRM for p>1.

The proposed estimator CMB possesses slightly worse accuracy compared to the MLE, GCM and MRM but it loses on the average not more than 50% (compared to the maximal MLE variance value). It is worth noting that CMB variance practically coincides with the variance values of MRM for p<0.6 and p>1.6 and sometimes can be even smaller (for p≤0.5). Remember that moment-based methods are much less computation demanding than the MLE or GCM estimators. The same statement is correct for the proposed CMB and TH.

5. CONCLUSIONS

It is shown that the best accuracy of GGD shape parameter estimation can be achieved by the maximum likelihood procedure and global convergence method. Both of them should be realized using iterative procedures and are rather computationally demanding. Almost the same accuracy can be obtained by the moment-based method MRM. Slightly greater variance values are observed for the proposed estimator CMB on the basis of the forth central moment and robust measure of kurtosis coefficient. The main advantage of the latter estimators is their relative simplicity from the computation point of view. The entropy-matching estimators proved to be biased for p>1 mainly because of the high steepness of entropic functions.

6. REFERENCES

[1] N. Ponomarenko, V. Lukin, S. Abramov, K. Egiazarian, J. Astola, "Blind evaluation of additive noise variance in textured images by nonlinear processing of block DCT coefficients," *Image Processing: Algorithms and Systems II: proc. of Internat. Conf.*, USA, 2003, SPIE Vol. 5014, pp. 178-189.

[2] A. Bovik, *Handbook on Image and Video Processing*, Academic Press, USA, 2000.

[3] M. Saad, A. Bovik, C. Charrier, "DCT-Statistics Model-Based Blind Image Quality Assessment," *Proceedings of ICIP*, 2011, 4 p.

[4] D. Zoran, Y. Weiss, "Scale Invariance and Noise in Natural Images," *Proceedings of ICCV*, 2009, pp. 2209-2216.

[5] F. Muller, "Distribution shape of two-dimensional DCT coefficients of natural images," *Electronic Letters* 29 (22), 1993, pp. 1935–1936.

[6] S. Mallat, "A theory of multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.* 11 (7), 1989, pp. 674–693.

[7] A. Briassouli, P. Tsakalides, A. Stouraitis, "Hidden Messages in Heavy Tails: DCT-Domain Watermark Detection Using Alpha-Stable Models," *IEEE Trans. Multimedia*, 7 (4), 2005, pp. 700-715.

[8] K. Kokkinakis, A. Nandi, "Exponent parameter estimation for generalized Gaussian probability density functions with application to speech modeling," *Signal Processing,* 85, 2005, pp. 1852-1858.

[9] M. Varanasi, B. Aazhang, "Parametric generalized Gaussian density estimation," *J. Ac. Soc. Amer.* 86 (4), 1989, pp. 1404-1415.

[10] A. Roenko, D. Kurkin, V. Lukin, "New estimators for tail heaviness parameter of generalized Gaussian distribution," *Mediterr. Conf. on Embed. Computing*, 2012, pp. 96-99

[11] R. Joshi, T. Fisher, "Comparison of generalized Gaussian and Laplacian modeling in DCT image coding," *IEEE Signal Proc. Lett.* 2 (5), 1995, pp. 81-82.

[12] S. Yu, A. Zhang, H. Li, "A review of estimating the shape parameter of generalized Gaussian distribution," *Journal of Comput. Inform. Systems*, 8: 21, 2012, pp. 9055-9064.

[13] A. Ralston, P. Rabinowitz, *First Course in Numerical Analysis*, McGraw-Hill, New York, 1978.

[14] M. Pi, "Improve maximum likelihood estimation for subband GGD parameters," *Pattern Recogn. Lett.* 27, 2006, pp. 1710-1713.

[15] K. Sharifi, A. Garcia, "Estimation of shape parameter for generalized Gaussian distribution in subband decompositions of video," *IEEE Transactions on Circuits and Systems for Video Technology*, 5, 1995, pp. 52-56.

[16] K. Song, "A globally convergent and consistent method for estimating the shape parameter of a generalized Gaussian distribution," *IEEE Trans. Inform. Theory* 52, 2006, pp. 510-527.

[17] B. Aiazzi, L. Alparone, S. Baronti, "Estimation based on entropy matching for generalized Gaussian PDF modeling," *IEEE Signal Processing Letters* 6, 1999, pp. 138-140.