Asymmetric Generalized Gaussian Distribution Parameters Estimation based on Maximum Likelihood, Moments and Entropy

Nafaa Nacereddine
Research Center in Industrial Technologies
P.O.Box 64, 16014 Algiers, Algeria
Email: n.nacereddine@crti.dz

Aicha Baya Goumeidane
Research Center in Industrial Technologies
P.O.Box 64, 16014 Algiers, Algeria
Email: a.goumeidane@crti.dz

***Abstract*—In this paper, we address the problem of estimating the parameters of Asymmetric Generalized Gaussian Distribution (AGGD) using three estimation mehods, namely, Maximum Likelihood Estimation (MLE), Moment Matching Estimation (MME) and Entropy Matching Estimaion (EME). For this purpose, these methods are applied on an unimodal histogram fitting of an image corrupted with AGGD noise. Experiments show that the effectiveness of each method comparatively to the other one depends on the variation range of the shape factor.**

***Index Terms*—Asymmetric generalized Gaussian distribution, Parameter estimation, Maximum likelihood, Moments, Entropy.**

I. Introduction

Although the major part of the statistical pattern recognition approaches considers models using the Gaussian distribution for their simplicity of computation, the modern systems of signal processing, compuer vision, image analysis, complex probability density function (*pdf*) approximation and communication, involving statistical modeling of data must often operate in complex and variable noise environments dominated by non-Gaussian sources [1], [2]. For example, a normalized histogram or *pdf* of a real image may be inherently non-Gaussian or even a mix of several distributions, so using a mixture of Gaussian distribution cannot portray the data accurately. As alternative, the use of an asymmetric generalized Gaussian distribution (AGGD) not only approximate a large class of statistical distributions (e.g. impulsive, Laplacian, Gaussian and uniform distributions) but also include the asymmetry [3]. The modelling capacity confers to this distribution more flexibility compared to the other distributions.

The main method used to estimate the statistical distribution is the maximum likelihood thanks to the development of the Expectation-Maximization algorithm by Dempster et el.(dempster) which is adapted for statistical data presented in form of mixture models. However, recently, other methods based on moments, L-moments, entropy, etc. have been also developed to estimate the distribution parameters for a large class of distributions such Generalized Gaussian distribution [2], Beta Laplace distribution [5] and Asymmetric Exponential Power distribution [6], [7]. Concerning the AGGD parameters estimation which is the subject of this paper, the authors in [3] have proposed the MLE method to achieve this estimation while, in [8], the authors have developed a moment-based method to estimate the AGGD parameters.

In this paper, the method Entropy Matching Estimation (EME) is newly studied and developed in case of an AGGD where, the parameters estimates are compared with methods of Maximum Likelihood Estimation (MLE) and Moment Matching Estimation (MME).

II. Definition and some properties

A. Asymmetric Generalized Gaussian distribution

We define the probability density function (pdf) of an unidimensional asymmetric generalized Gaussian distribution $(AGGD)$ as follows

$$f(x|\boldsymbol{\theta}) = \begin{cases} \dfrac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} e^{-[(-x+\mu)/\alpha_1]^\beta} & \text{if } x < \mu \\ \dfrac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} e^{-[(x-\mu)/\alpha_2]^\beta} & \text{if } x \geq \mu \end{cases} \tag{1}$$

for x ($x \in \mathbb{R}$), and where $\boldsymbol{\theta} = (\mu, \alpha_1, \alpha_2, \beta)^T$ ($-\infty < \mu < \infty, \alpha_1 > 0, \alpha_2 > 0, \beta > 0$) is the distribution parameter vector whose components are the left and right scale parameters and shape parameter, respectively. Γ is the gamma function defined by $\Gamma(\xi) = \int_0^\infty e^{-t} t^{\xi-1}\, \mathrm{d}t$. The parameter β dictates the exponential rate of decay: the larger the values of β the flatter the *pdf*; the smaller the β the more peaked the *pdf*, as shown in Fig. 1.

Special cases of AGGD can be obtained with some values of β as follows : asymmetric Laplacian distribution for $\beta = 1$, asymmetric Gaussian distribution for $\beta = 2$, impulse distribution for $\beta \to 0$ and uniform distribution for $\beta \to \infty$. Note that the scale parameters which express the width of the distribution are linked to standard deviations by

$$\alpha_i = \sigma_i \sqrt{\frac{\Gamma(1/\beta)}{\Gamma(3/\beta)}}, \qquad i = 1, 2 \tag{2}$$

978-1-7281-4914-1/19/$31.00 2019 IEEE

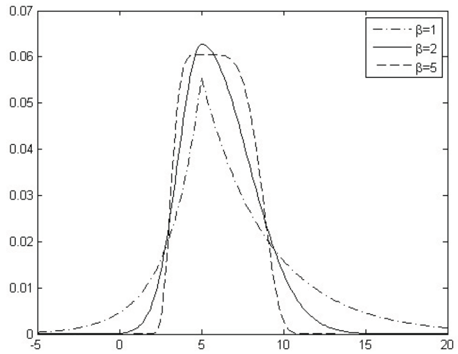

Fig. 1. AGGD's *pdf*s for $\mu = 5$, $\alpha_1 = 2$, $\alpha_2 = 4$, $\beta = \{1, 2, 5\}$ [8]

B. Cumulative distribution function

Proposition 1. Let X be an AGGD random variable with *pdf* f as defined in (1), then for any $x \in \mathbb{R}$, the cumulative distribution function (*cdf*) F can be expressed by :

$$F(x|\boldsymbol{\theta}) = \begin{cases} \dfrac{\alpha_1}{(\alpha_1+\alpha_2)\Gamma(1/\beta)}\,\Gamma\left[\dfrac{1}{\beta}, \left(\dfrac{-x+\mu}{\alpha_1}\right)^{\beta}\right] & \text{if } x < \mu \\ \dfrac{1}{\alpha_1+\alpha_2}\left[\alpha_1 + \dfrac{\alpha_2}{\Gamma(1/\beta)}\,\gamma\left[\dfrac{1}{\beta}, \left(\dfrac{x-\mu}{\alpha_2}\right)^{\beta}\right]\right] & \text{if } x \geq \mu \end{cases} \quad (3)$$

where $\gamma(\cdot\,,\cdot)$ and $\Gamma(\cdot\,,\cdot)$ denote respectively the lower and the upper incomplete gamma functions. They are defined as $\gamma(s,y) = \int_0^y t^{s-1}e^{-t}\,\mathrm{d}t$ and $\Gamma(s,y) = \int_y^\infty t^{s-1}e^{-t}\,\mathrm{d}t$, $y \geq 0$.

Proof. See Appendix A □

C. Moments

Some of the most important features and charcteristics of a distribution can be studied through moments (e.g. tendency, dispersion, skewness and kurtosis) [5].

Proposition 2. If X is a random variable with the AGGD density and, for any $r \in \mathbb{N}$, the r-order non central moment can be written as

$$\mathbb{E}[X^r] = \frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} \times \sum_{k=0}^{r} \binom{r}{k} \left[(-1)^k \alpha_1^{k+1} + \alpha_2^{k+1}\right] \Gamma\left(\frac{k+1}{\beta}\right) \mu^{r-k} \quad (4)$$

Proof. See [8]. □

Proposition 3. Using the proposition 1, the first four moments (mean, variance, skewness, and kurtosis) of X are deduced as

1) Mean

$$me = \mu - (\alpha_1 - \alpha_2)\frac{\Gamma(2/\beta)}{\Gamma(1/\beta)} \quad (5)$$

2) Variance

$$\sigma^2 = \frac{\alpha_1^3+\alpha_2^3}{\alpha_1+\alpha_2} \times \frac{\Gamma(3/\beta)}{\Gamma(1/\beta)} - (\alpha_1-\alpha_2)^2 \frac{\Gamma(2/\beta)^2}{\Gamma(1/\beta)^2} \quad (6)$$

3) Skewness

$$\gamma_1 = \frac{\alpha_1-\alpha_2}{\sigma^3} \times \left[3\frac{\alpha_1^3+\alpha_2^3}{\alpha_1+\alpha_2}\frac{\Gamma(2/\beta)\Gamma(3/\beta)}{\Gamma(1/\beta)^2} - 2(\alpha_1-\alpha_2)^2\frac{\Gamma(2/\beta)^3}{\Gamma(1/\beta)^3} - (\alpha_1^2+\alpha_2^2)\frac{\Gamma(4/\beta)}{\Gamma(1/\beta)}\right] \quad (7)$$

4) Kurtosis

$$\gamma_2 = \frac{1}{\sigma^4} \times \left[\frac{\alpha_1^5+\alpha_2^5}{\alpha_1+\alpha_2} \times \frac{\Gamma(5/\beta)}{\Gamma(1/\beta)} - (\alpha_1-\alpha_2)^2 \left(4(\alpha_1^2+\alpha_2^2)\frac{\Gamma(2/\beta)\Gamma(4/\beta)}{\Gamma(1/\beta)^2} - 6\frac{\alpha_1^3+\alpha_2^3}{\alpha_1+\alpha_2} \times \frac{\Gamma(2/\beta)^2\Gamma(3/\beta)}{\Gamma(1/\beta)^3} + 3(\alpha_1-\alpha_2)^2\frac{\Gamma(2/\beta)^4}{\Gamma(1/\beta)^4}\right)\right] - 3 \quad (8)$$

For Skewness and Kurtosis, σ is expressed in (6).

Proof. See Appendix B □

D. Entropy

The entropy is the basic concept in Information Theory [9] which measures the uncertainty associated to a random process. For a discrete random variable X with $x_1, x_2, \ldots, x_n$ the set of possible values, the entropy is given as

$$\begin{aligned} H(X) &\equiv -\mathbb{E}[\log f(X)] \\ &= -\textstyle\sum_{i=1}^{i=n} f(X=x_i)\log f(X=x_i) \end{aligned} \quad (9)$$

For continuous random variables, it is called differential entropy.

Proposition 4. The differential entropy of a continuous random variable X following an AGGD with parameters $\boldsymbol{\theta}$, is computed, in *bits*, as

$$\begin{aligned} H(\boldsymbol{\theta}) &= -\textstyle\int_{-\infty}^{+\infty} f(x,\boldsymbol{\theta})\log_2\left(f(x,\boldsymbol{\theta})\right)\mathrm{d}x \\ &= \tfrac{1}{\beta\log 2} - \log_2 \tfrac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} \end{aligned} \quad (10)$$

Proof. See Appendix C □

III. AGGD PARAMETERS ESTIMATION

The estimation of the AGGD parameters efficiently is necessary to allow this distribution to better fit the raw signal sources. For this purpose, three estimation methods based on the likelihood function, the moments and the entropy are developed in the next subsections.

A. Maximum Likelihood Estimation

For a sample $x_1, x_2, \ldots, x_n$ of n iid[1] observations, the likelihood function is given by

$$\mathcal{L}(\boldsymbol{\theta}|x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n|\boldsymbol{\theta}) \overset{iid}{=} \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) \quad (11)$$

[1] independent and identically distributed

In practice it is often more convenient to work with the logarithm of the likelihood function, i.e.

$$\log \mathcal{L}(\boldsymbol{\theta}|x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \log f(x_i|\boldsymbol{\theta}) \qquad (12)$$

To simplify, $\log \mathcal{L}(\boldsymbol{\theta}|x_1, x_2, \ldots, x_n)$ is denoted $\mathcal{L}(\boldsymbol{\theta}|x)$. Thus, for AGGD random variable, we have

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}|x) = & \sum_{x_i<\mu} \log \left[\tfrac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} e^{-[(-x_i+\mu)/\alpha_1]^{\beta}}\right] \\ & + \sum_{x_i\geq\mu} \log \left[\tfrac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} e^{-[(x_i-\mu)/\alpha_2]^{\beta}}\right] \\ = & \; n\log\beta - n\log(\alpha_1+\alpha_2) - n\log\Gamma(1/\beta) \\ & - \sum_{x_i<\mu}\left(\tfrac{-x_i+\mu}{\alpha_1}\right)^{\beta} - \sum_{x_i\geq\mu}\left(\tfrac{x_i-\mu}{\alpha_2}\right)^{\beta}\end{aligned} \qquad (13)$$

The maximum likelihood estimate (MLE) $\hat{\boldsymbol{\theta}}_{mle} = (\hat{\mu}_{mle}, \hat{\alpha}_{1_{mle}}, \hat{\alpha}_{2_{mle}}, \hat{\beta}_{mle})^T$ of $\boldsymbol{\theta}$, defined in §2.1, is obtained by maximizing the log-likelihood function $\mathcal{L}(\boldsymbol{\theta}|x)$ w.r.t. the unknown parameter vector $\boldsymbol{\theta}$. The ML solution for each unknown parameter θ_j, $(j = 1, \ldots, 4)$ can be found by setting the partial derivative of $\mathcal{L}(\boldsymbol{\theta}|x)$ with regard to θ_j equal to zero, assuming that the parameters θ_k $(k \neq j)$ are known, yielding the following likelihood equations:

$$\frac{1}{\alpha_1^{\beta}} \sum_{x_i<\hat{\mu}_{mle}} (-x_i + \hat{\mu}_{mle})^{\beta-1} - \frac{1}{\alpha_2^{\beta}} \sum_{x_i\geq\hat{\mu}_{mle}} (x_i - \hat{\mu}_{mle})^{\beta-1} = 0 \qquad (14)$$

$$n\hat{\alpha}_{1_{mle}}^{\beta+1} - \beta(\hat{\alpha}_{1_{mle}} + \alpha_2) \sum_{x_i<\mu} (-x_i + \mu)^{\beta} = 0 \qquad (15)$$

$$n\hat{\alpha}_{2_{mle}}^{\beta+1} - \beta(\hat{\alpha}_{2_{mle}} + \alpha_1) \sum_{x_i\geq\mu} (x_i - \mu)^{\beta} = 0 \qquad (16)$$

$$\begin{aligned}\frac{1}{\hat{\beta}_{mle}} + \frac{\Psi(1/\hat{\beta}_{mle})}{\hat{\beta}_{mle}^2} - \sum_{x_i<\mu}\left(\frac{-x_i+\mu}{\alpha_1}\right)^{\hat{\beta}_{mle}} \log\left(\frac{-x_i+\mu}{\alpha_1}\right) \\ - \sum_{x_i\geq\mu}\left(\frac{x_i-\mu}{\alpha_2}\right)^{\hat{\beta}_{mle}} \log\left(\frac{x_i-\mu}{\alpha_2}\right) = 0\end{aligned} \qquad (17)$$

with $\Psi(\xi) = \frac{\mathrm{d}\log\Gamma(\xi)}{\mathrm{d}\xi}$, the digamma function

These equations can be solved numerically via iterative methods such as Newton-Raphson method.

B. Moment Matching Estimation

Knowing that $\{x_1, x_2, \ldots, x_n\}$ is a set of n realizations of the random variable X, the empirical moments or sample moments denoted $me_s, \sigma_s^2, \gamma_{1s}$ and γ_{2s} are defined by

$$me_s = \bar{x} = \sum_{i=1}^{n} x_i/n \qquad (18)$$

$$\sigma_s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/n \qquad (19)$$

$$\gamma_{1s} = \sum_{i=1}^{n} (x_i - \bar{x})^3/(n\sigma_s^3) \qquad (20)$$

$$\gamma_{2s} = \sum_{i=1}^{n} (x_i - \bar{x})^4/(n\sigma_s^4) \qquad (21)$$

The moment method estimation (MME) may be accomplished by equating the theoretical expressions of moments me, σ^2, γ_1 and γ_2 given by equations (5)-(8) and samples moment values given in (18)-(21) and then solving numerically the nonlinear system to obtain the estimated parameter vector $\boldsymbol{\theta}_{mme} = (\hat{\mu}_{mme}, \hat{\alpha}_{1_{mme}}, \hat{\alpha}_{2_{mme}}, \hat{\beta}_{mme})^T$.

C. Entropy Matching Estimation

This class of estimators relies on matching the entropy of a distribution as a source model with that of the empirical source data. Here, such a distribution is the AGGD of which the entropy will be fitted to the source signal. Given an AGGD, its differential entropy $H(\boldsymbol{\theta})$, given by (10), is only a function of σ_1, σ_2 and β without being dependent of μ because of entropy invariance to translation. Suppose that $H(X)$, defined in (9) and renamed H_X, is the zero-order entropy of the data $x_1, x_2, \ldots, x_n$ obtained at the output of a uniform threshold quantizer with step size Δ. In [10], Woods has established that $H(\theta)$ (called here H_{AGGD}) is linked with $H(X)$ by

$$H_X = H_{AGGD} - \log_2 \Delta \qquad (22)$$

using (2) and (10) leads to

$$H_X - \log_2 \frac{\sigma_1 + \sigma_2}{\Delta} = \frac{1}{\beta \log 2} - \log_2 \frac{\beta\Gamma(3/\beta)^{1/2}}{\Gamma(1/\beta)^{3/2}} \triangleq \mathcal{F}_H(\beta) \qquad (23)$$

By plotting $\mathcal{F}_H$ in function of β, it is noted that this function is increasing in the interval $]0 \;\; 2]$ where, it reaches its maximum (equal to 2.047) for $\beta = 2$, as illustrated in Fig. 2.

Knowing the values of Δ, σ_{1s}, σ_{2s} and H_X which are computed from the real data, the estimated value of the shape factor $\hat{\beta}_{eme}$ is derived by inverting $\mathcal{F}_H$, i.e.,

$$\hat{\beta}_{eme} = \mathcal{F}_H^{-1}\left(H_X - \log_2 \frac{\sigma_{1s} + \sigma_{2s}}{\Delta}\right) \qquad (24)$$

If the data consists in an histogram of which the gray level values belong to $\{0, 1, \ldots, 255\}$, then Δ is taken equal to 1. About the computation of the empirical left and right standard deviations, using the AGGD analytical expression, a first approximation consists to express them in function of the whole sample variance σ_s^2 of AGGD as follows

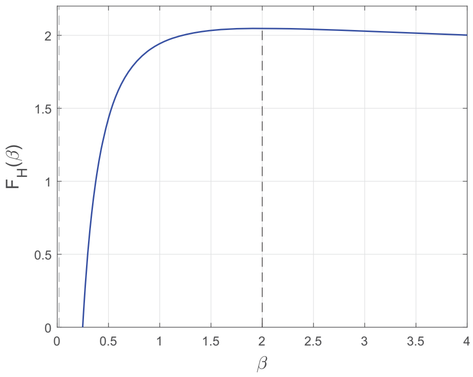

Fig. 2. Entropy-based AGGD shape function

$$\sigma_{1s} + \sigma_{2s} \sim 2\sqrt{\sigma_s^2} = 2\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \sum_{i=1}^{n} x_i/n\right)^2} \quad (25)$$

However, here, based on the analytical expression of AGGD which may be considered as the union between the left and right side of two GGDs with different variances as illustrated in Fig. 3, we have computed the AGGD left and right sample variances σ_{1s} and σ_{2s} respectively, as follows

$$\sigma_{1s}^2 = \sum_{i=1}^{i_{x_m}} x_i \left(i - i_{x_m}\right)^2 \bigg/ \sum_{i=1}^{i_{x_m}} x_i \quad (26)$$

and

$$\sigma_{2s}^2 = \sum_{i=i_{x_m}+1}^{n} x_i \left(i - i_{x_m}\right)^2 \bigg/ \sum_{i=i_{x_m}+1}^{n} x_i \quad (27)$$

where

$$i_{x_m} = \underset{i=1,\ldots,n}{\operatorname{argmax}}\, x_i \quad (28)$$

Using (2), the estimate scale parameters for EME are deduced as

$$\hat{\alpha}_{i_{eme}} = \sigma_{is}\sqrt{\frac{\Gamma(1/\hat{\beta}_{eme})}{\Gamma(3/\hat{\beta}_{eme})}}, \qquad i = 1, 2 \quad (29)$$

IV. Experiments

Let $h_g(g \in [0, L-1]$ where L is the number of gray levels) the normalized histogram of an image X which can be seen as an estimate of the true $pdf\, f(g)$ of the image. In order to assess the effectiveness of the proposed parameter estimations, namely, Maximum Likelihood Estimation (MLE), Moment Matching Estimation (MME) and Entropy Matching Estimation (EME) for unimodal histogram fitting, a simulated image corrupted by asymmetric generalized Gaussian noise with single mean and scales parameters (μ, α_1, α_2) and different shape parameters (β) is used. In our experiments, we take $\mu = 100$, $\alpha_1 = 15$ and $\alpha_2 = 30$. However, the shape factor β is varied from $\beta = 0$ to $\beta = 3$ with a step of $\Delta\beta = 0.2$. The synthetic image and its fitted histograms based

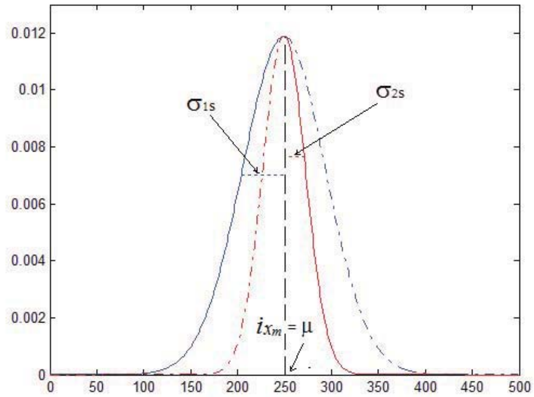

Fig. 3. **a.**(Blue and red solid lines) Left and right parts of AGGD. **b.**(Blue and red dashed lines) Symmetrical parts of (a.) w.r.t the vertical line passing through the mean μ. **c.**(Blue and red dotted lines) Left and right standard deviations for AGGD

on MLE, MME and EME are illustrated in Fig. 4 and Fig.5 and the resulted estimated AGGD parameters are reported in Table 1. The boldfaced value in Table 1 refers to the best result compared to the input data between the three studied parameter estimation methods, while the underlining is used to mark the value comparable to the best result. The value followed of an asterisk is dealing with erroneous values as we will explain later. From the simulated image, we try to show the strength and limitation of each of the AGGD parameter estimation methods, especially, the AGGD shape factor, by varying the values of β to display AGGDs from very peaked one ($\beta = 0.6$) to somewhat flatened one ($\beta = 3$) passing through Laplacian ($\beta = 1$) and ($\beta = 2$) Gaussian ones.

As it can be seen in Fig. 4, the best histogram fitting outcomes for low values of β are provided by the EME method which, according to the results given in Table 1, remains the leader until $\beta = 1.4$ with a comparable result with MLE is obtained for $\beta = 1.6$. Beyond this shape factor amount, EME method falls drastically while the method of MLE outperforms MME for $\beta = 1.8$. According to Fig. 5 and still according to Table 1, the results of MLE and MME methods are in a reasonable agreement with the input data from $\beta = 2$ to $\beta = 2.4$. However, the MME method slightly outperforms MLE method for the values of β greater than 2.4.

Since the shape factor plays a major role in the design of asymmetric generalized Gaussian distribution, the estimation of the AGGD scale parameters α_1 and α_1 is governed by the computation accuracy of the shape factor estimate. That is said, the AGGD scale parameters estimates provided in our experiments are in general in accordance with the outcomes of the shape factor. However, we remark that, for β greater than 2, the lower and the higher scale parameters α_1 and α_2 are better estimated by MLE and MME, respectively.

For the AGGD mean parameter μ, as mentioned in §III.C, the entropy is invariant to translation so, here, the AGGD μ parameter is computed empirically and which corresponds to

		MLE	MME	EME
β=0.6	$\hat{\mu}$	98.89	101.27	**99**
	$(\hat{\alpha_1}, \hat{\alpha_2})$	(30.47 , 42.44)	(41.1 , 52.76)	**(24.98 , 31.38)**
	$\hat{\beta}$	1.27	1.71	**1.05**
β=0.8	$\hat{\mu}$	100.2	99.9	**100**
	$(\hat{\alpha_1}, \hat{\alpha_2})$	(20.86 , 30.27)	(28.41 , 40.93)	**(19.47 , 26.48)**
	$\hat{\beta}$	1.11	1.46	**1.04**
β=1	$\hat{\mu}$	99.38	100.99	**100**
	$(\hat{\alpha_1}, \hat{\alpha_2})$	(16.51 , **26.58)**	(21.9 , 31.4)	(**16.28 ,** 23.02)
	$\hat{\beta}$	1.12	1.35	**1.06**
β=1.2	$\hat{\mu}$	100.98	98.03	**100**
	$(\hat{\alpha_1}, \hat{\alpha_2})$	(17.01 , **25.68)**	(17.12 , 31.16)	(**15.93 ,** 23.87)
	$\hat{\beta}$	1.28	1.46	**1.21**
β=1.4	$\hat{\mu}$	100.89	98.94	**100**
	$(\hat{\alpha_1}, \hat{\alpha_2})$	(16.58 , **25.16)**	(16.26 , 28.44)	(**15.11 ,** 23.05)
	$\hat{\beta}$	1.48	1.61	**1.35**
β=1.6	$\hat{\mu}$	**99.67**	99.39	99
	$(\hat{\alpha_1}, \hat{\alpha_2})$	**(15.33 , 26.41)**	(16.42 , 27.43)	(14.39 , 23.65)
	$\hat{\beta}$	**1.69**	1.82	1.5
β=1.8	$\hat{\mu}$	99.48	**99.85**	99
	$(\hat{\alpha_1}, \hat{\alpha_2})$	**(15.09 , 26.40)**	(16.16 , 26.45)	(13.81 , 23.09)
	$\hat{\beta}$	**1.83**	1.91	1.55
β=2	$\hat{\mu}$	100.38	100.26	**100**
	$(\hat{\alpha_1}, \hat{\alpha_2})$	(**15.74 ,** 25.87)	(16.14 , **25.78)**	(13.57 , 21.67)
	$\hat{\beta}$	**2.04**	2.06	1.57
β=2.2	$\hat{\mu}$	98.45	**100.7**	98
	$(\hat{\alpha_1}, \hat{\alpha_2})$	(**13.98 ,** 27.42)	(16.3 , **25.27)**	(11.47*, 21.36*)
	$\hat{\beta}$	**2.24**	**2.24**	1.53*
β=2.4	$\hat{\mu}$	99.04	**100.7**	99
	$(\hat{\alpha_1}, \hat{\alpha_2})$	(**15.07 ,** 26.75)	(16.73 , **25.28)**	(11.83*, 20.03*)
	$\hat{\beta}$	2.44	**2.43**	1.53*
β=2.6	$\hat{\mu}$	**99.45**	101.04	99
	$(\hat{\alpha_1}, \hat{\alpha_2})$	(**15.09 ,** 27.18)	(17.14 , **25.21)**	(11.01*, 19.22*)
	$\hat{\beta}$	2.71	**2.66**	1.45*
β=2.8	$\hat{\mu}$	101.21	**100.95**	101
	$(\hat{\alpha_1}, \hat{\alpha_2})$	**(16.82 , 24.83)**	(16.95 , 24.6)	(11.54*, 16.55*)
	$\hat{\beta}$	2.85	**2.80**	1.4*
β=3	$\hat{\mu}$	**99.22**	101.39	99
	$(\hat{\alpha_1}, \hat{\alpha_2})$	(**15.26 ,** 26.85)	(17.47 , **24.61)**	(9.93*, 16.88*)
	$\hat{\beta}$	3.14	**3.09**	1.35*

TABLE I
MLE, MME AND EME RESULTS FOR AGGD WITH INPUT PARAMETERS μ=100, α_1=15, α_2=25 AND β=0.6:0.2:3

the maximum of the distribution, i.e., $\mu_s = \text{argmax}_g\, h_g$. The mean parameter is also delicate to compute for MLE because of its highly non linear expression (see (14)) handling, among others, piece-wise functions. For EME, it is expected that the values of the shape factor estimates for the input β greater than 2 are biased since the the function $\mathcal{F}_H(\beta)$ given in (23) is non-invertible and is only valid for β less that 2. Consequently, in this case, the related scale parameters of AGGD are also biased since they are linked to β by (29).

It is worth to note, in terms of computational time, that EME method is much more faster than MLE and MME methods where the EME average running time is around 0.1 s against 0.6 s and 1.3 s, for MLE and MME, respectively. This quality makes the utilization of EME more suitable for real-time signal processing applications with the condition that the fitted signals present low shape factor values, as depicted above.

V. CONCLUSION

In this paper, the gray level histogram of a an unimodal synthetic image corrupted by asymmetric generalized Gaussian noise with different shape factor values is used in order to assess the effectiveness of the methods of Maximum Likelihood Estimation (MLE), Moment Matching Estimation (MME) and Entropy Matching Estimation (EME) for the estimation of the AGGD noise parameters. The histogram fitting results for each estimation method are dependent on the value of the shape factor β where, EME is more suitable for low shape factor values (less than 1.6) while the best estimates are obtained by the MLE and the MME methods for $\beta > 1.6$ with slight advantages to MLE for $\beta < 1.8$ and MME for $\beta > 2.4$. It is also noticed that the EME method is computationally much more faster than the two others which makes it predisposed to real-time utilization in the data modelling area.

APPENDIX

A. *Proof of Proposition 1.*

For an AGGD, the *cdf* is defined as follows

$$F(x) = \begin{cases} \dfrac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} \displaystyle\int_{-\infty}^{x} e^{-[(-y+\mu)/\alpha_1]^\beta}\, \mathrm{d}y & \text{if } x < \mu \\ \dfrac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} \left[\displaystyle\int_{-\infty}^{\mu} e^{-[(-y+\mu)/\alpha_1]^\beta}\, \mathrm{d}y + \int_{-\mu}^{x} e^{-[(y-\mu)/\alpha_2]^\beta}\, \mathrm{d}y\right] & \text{if } x \geq \mu \end{cases} \tag{30}$$

Performing the changes of variable $t = [(-y+\mu)/\alpha_1]^\beta$ and $u = \frac{-x+\mu}{\alpha_1}$ in the equation part taken for $x < \mu$, leads to the following equalities:

$$\begin{aligned} F(x) &= \frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} \frac{\alpha_1}{\beta} \int_{u^\beta}^{\infty} t^{1/\beta-1} e^{-t}\, \mathrm{d}t \\ &= \frac{\alpha_1}{(\alpha_1+\alpha_2)\Gamma(1/\beta)}\, \Gamma\left[\frac{1}{\beta}, \left(\frac{-x+\mu}{\alpha_1}\right)^\beta\right] \qquad \text{for } x < \mu \end{aligned}$$

For $x \geq \mu$, taking the change of variable $t = [(-y+\mu)/\alpha_1]^\beta$, the term $\int_{-\infty}^{\mu} e^{-[(-y+\mu)/\alpha_1]^\beta}\, \mathrm{d}y$ in (30) could be replaced by $\frac{\alpha_1}{\beta} \int_{u}^{\infty} t^{1/\beta-1} e^{-t}\, \mathrm{d}t$ which is equal to $\frac{\alpha_1}{\beta}\Gamma(1/\beta)$. Otherwise, using the variable changes $t = [(y-\mu)/\alpha_2]^\beta$ and $u = \frac{x-\mu}{\alpha_2}$ the term $\int_{-\mu}^{x} e^{-[(y-\mu)/\alpha_2]^\beta}\, \mathrm{d}y$ in (30) becomes $\frac{\alpha_2}{\beta} \int_{0}^{u^\beta} t^{1/\beta-1} e^{-t}\, \mathrm{d}t$ which is equal to $\frac{\alpha_2}{\beta}\gamma\left[\frac{1}{\beta}, \left(\frac{x-\mu}{\alpha_2}\right)^\beta\right]$.

Then, we obtain

$$F(x) = \frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)}\left[\frac{\alpha_1}{\beta}\Gamma(1/\beta) + \frac{\alpha_2}{\beta}\gamma\left[\frac{1}{\beta}, \left(\frac{x-\mu}{\alpha_2}\right)^{\beta}\right]\right]$$
$$= \frac{\alpha_1}{(\alpha_1+\alpha_2)} + \frac{\alpha_2}{(\alpha_1+\alpha_2)\Gamma(1/\beta)}\gamma\left[\frac{1}{\beta}, \left(\frac{x-\mu}{\alpha_2}\right)^{\beta}\right] \quad \text{for } x \geq \mu$$

B. Proof of Proposition 3.

The four moments are obtained using Proposition 2 and the following definitions of mean (me), variance (σ^2), skewness (γ_1) and kurtosis (γ_2)

$$\begin{cases} me &= \mathbb{E}[X] \\ \sigma^2 &= \mathbb{E}[X-\mathbb{E}[X]]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ \gamma_1 &= \frac{\mathbb{E}[X-\mathbb{E}[X]]^3}{\sigma^3} = \frac{\mathbb{E}[X^3]-3\mathbb{E}[X]\mathbb{E}[X^2]+2\mathbb{E}[X]^3}{\sigma^3} \\ \gamma_2 &= \frac{\mathbb{E}[X-\mathbb{E}[X]]^4}{\sigma^4} - 3 \\ &= \frac{\mathbb{E}[X]^4-4\mathbb{E}[X]\mathbb{E}[X^3]+6\mathbb{E}[X]^2\mathbb{E}[X^2]-3\mathbb{E}[X]^4}{\sigma^4} - 3 \end{cases}$$

C. Proof of Proposition 4.

For the sake of simplicity, we express first the entropy in *nats* and then we give at the end, its expression in *bits*

$$H(\boldsymbol{\theta}) = -\int_{-\infty}^{+\infty} f(x,\boldsymbol{\theta})\log\left(f(x,\boldsymbol{\theta})\right)\mathrm{d}x$$
$$= -\int_{-\infty}^{\mu} \frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} e^{-[(-x+\mu)/\alpha_1]^{\beta}} \times$$
$$\log\left[\frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} e^{-[(-x+\mu)/\alpha_1]^{\beta}}\right]\mathrm{d}x$$
$$-\int_{\mu}^{\infty} \frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} e^{-[(x-\mu)/\alpha_2]^{\beta}} \times$$
$$\log\left[\frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} e^{-[(x-\mu)/\alpha_2]^{\beta}}\right]\mathrm{d}x$$
$$= -(H_1 + H_2)$$

where

$$H_1 = \frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} \times$$
$$\left[\log\frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)}\int_{-\infty}^{\mu} e^{-[(-x+\mu)/\alpha_1]^{\beta}}\,\mathrm{d}x - \int_{-\infty}^{\mu}\left(\frac{-x+\mu}{\alpha_1}\right)^{\beta} e^{-[(-x+\mu)/\alpha_1]^{\beta}}\,\mathrm{d}x\right]$$

By using the change of variable $u = \frac{-x+\mu}{\alpha_1}$, we obtain

$$H_1 = \frac{\alpha_1\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} \times$$
$$\left[-\log\frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)}\int_0^{\infty} e^{-u^{\beta}}\,\mathrm{d}u + \int_0^{\infty} u^{\beta}e^{-u^{\beta}}\,\mathrm{d}u\right]$$
$$= \frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} \times$$
$$\left[\frac{\alpha_1}{\beta}\Gamma\left(\frac{1}{\beta}\right)\log\frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} - \frac{\alpha_1}{\beta}\Gamma\left(\frac{\beta+1}{\beta}\right)\right]$$
$$= \frac{\alpha_1}{\alpha_1+\alpha_2}\left[\log\frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} - \frac{1}{\beta}\right]$$

since $\Gamma\left(\frac{\beta+1}{\beta}\right) = \Gamma\left(\frac{1}{\beta}+1\right) = \frac{1}{\beta}\Gamma(1/\beta)$.

The same reasoning is followed for H_2 to obtain

$$H_2 = \frac{\alpha_2}{\alpha_1+\alpha_2}\left[\log\frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} - \frac{1}{\beta}\right]$$

The entropy H in *nats* is then expressed by

$$H = -(H_1 + H_2)$$
$$= \left[\frac{1}{\beta} - \log\frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)}\right]\left[\frac{\alpha_1}{\alpha_1+\alpha_2} + \frac{\alpha_2}{\alpha_1+\alpha_2}\right]$$
$$= \frac{1}{\beta} - \log\frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)}$$

Finally, the expression of the entropy in *bits* is given by

$$H = \frac{1}{\beta\log 2} - \log_2\frac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)}$$

References

[1] A. P. Benavent, F. E. Ruiz, and J. M. Sezaiazzi, "Learning Gaussian Mixture Models With Entropy-Based Criteria", *IEEE Trans. on Neural Networks* vol. 20, no. 11, pp. 1756–1771, 2009.

[2] B. Aiazzi, L. Alparone, and S. Baronti, "Estimation Based on Entropy Matching for Generalized Gaussian PDF Modeling", *IEEE Signal Processing Letters*, vol. 6, no. 6, pp. 138–140, 1999.

[3] N. Nacereddine, S. Tabbone, D. Ziou, L. Hamami, "Asymmetric generalized Gaussian mixture models and EM algorithm for image segmentation", *in Proc. 20th Int. Conf. Pattern Recognition (ICPR)*, pp. 4557–4560, Aug. 2010.

[4] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[5] G. M. Cordeiro, A. J. Lemonte, "The beta Laplace distribution", *Statistics & Probability Letters*, vol. 81, no. 8, pp. 973–982, 2011.

[6] P. Delicado, and M. N. Goria, "A small sample comparison of maximum likelihood, moments and L-moments methods for the asymmetric exponential power distribution", *Computational Statistics & Data Analysis*, vol. 52, no. 3, pp. 1661–1673, 2008.

[7] W. H. Asquith, "Parameter estimation for the 4-parameter Asymmetric Exponential Power distribution by the method of L-moments using R", *Computational Statistics and Data Analysis*, vol. 71, no. C, pp. 955–970, 2014.

[8] N. Nacereddine, D. Ziou, "Asymmetric Generalized Gaussian Mixtures for Radiographic Image Segmentation", *In Proc. of the 9th Int. Conf. on Computer Recognition Systems, CORES 2015*, AISC, vol. 403, Springer, pp. 521–532, 2016.

[9] T. M. Cover, and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, 2012.

[10] R. C Woods, "On optimum Quantization", *IEEE Trans. on Information Theory*, vol. 15, no. 2, pp. 248–252, 1969.

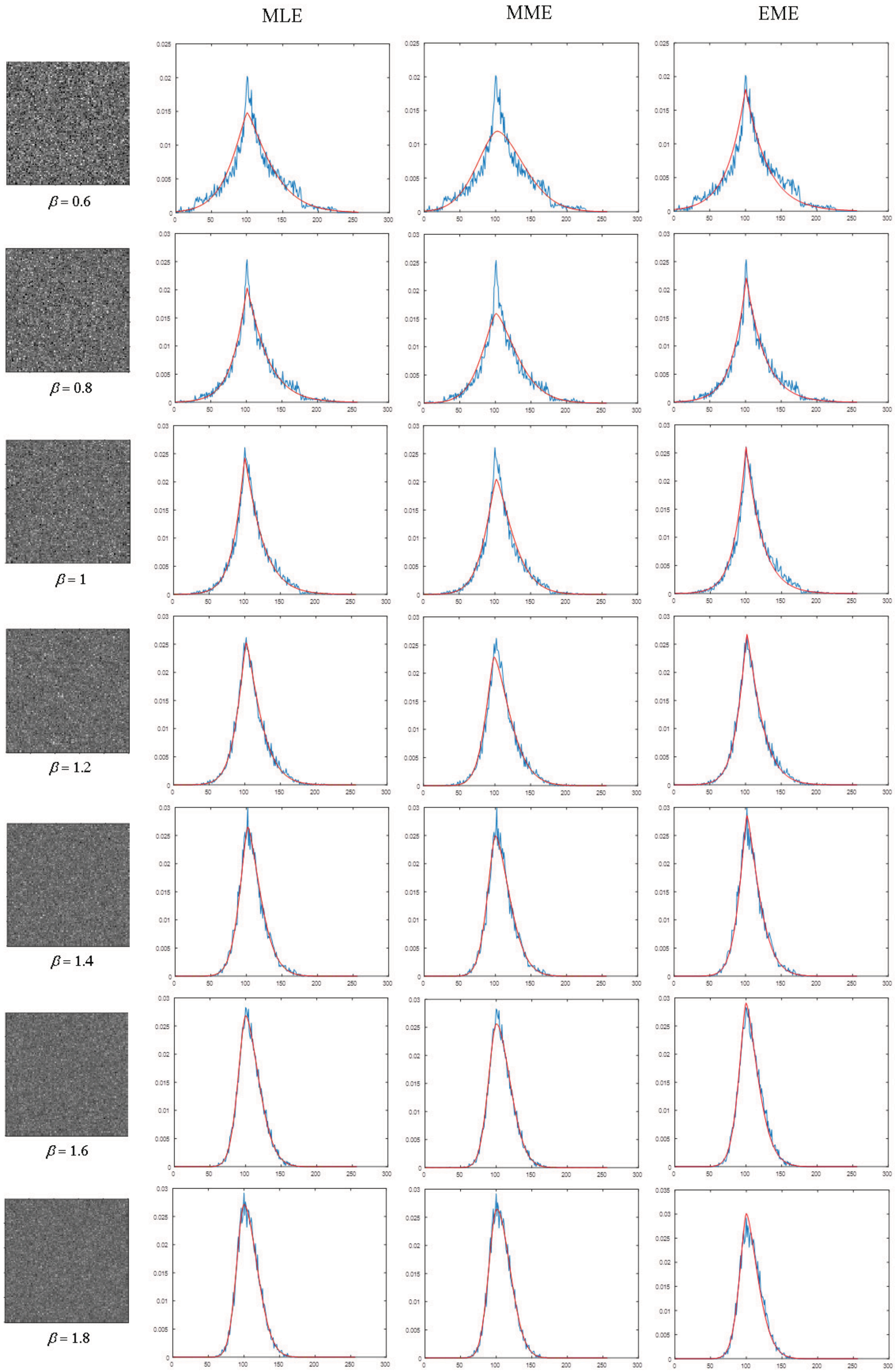

Fig. 4. From left to right: Image with gray level equal to $\mu = 100$ corrupted with AGGNoise para$(\alpha_1, \alpha_2) = (15, 25)$ and β, MLE-based, MME-based and EME-based histogram fitting. From top to bottom : the results of histogram fitting for various values of $\beta = \{0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2\}$ (Blue and red curves are original and fitted histograms, respectively)

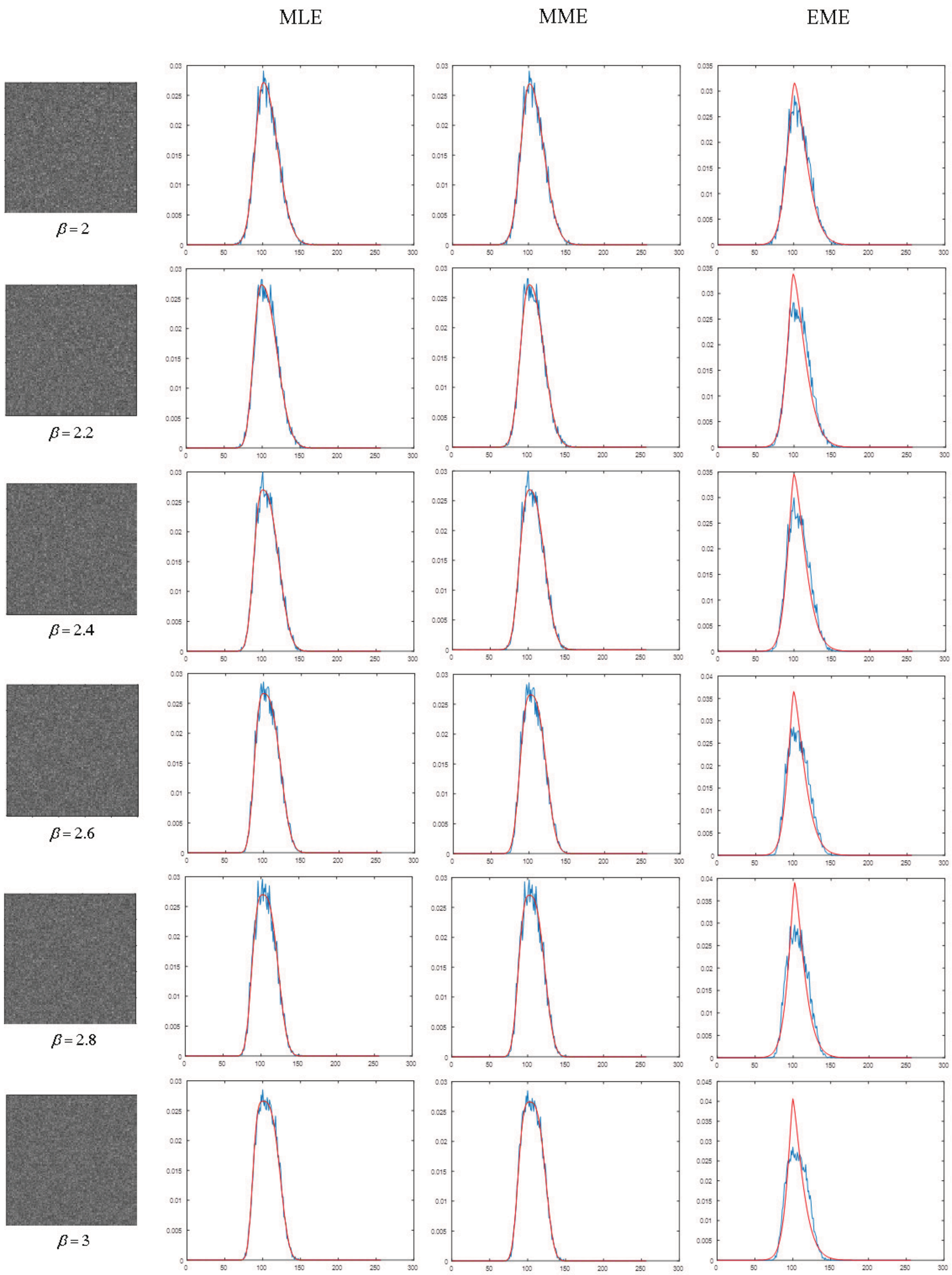

Fig. 5. From left to right: Image with gray level equal to $\mu = 100$ corrupted with AGGNoise$(\alpha_1, \alpha_2) = (15, 25)$ and β, MLE-based, MME-based and EME-based histogram fitting. From top to bottom : the results of histogram fitting for input values of $\beta = \{2, 2.2, 2.4, 2.6, 2.8, 3\}$ (Blue and red curves are original and fitted histograms, respectively)