Pattern Recognition 47 (2014) 1330–1348

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A generalized multiclass histogram thresholding approach based on mixture modelling

Aïssa Boulmerka [a], Mohand Saïd Allili [b,*], Samy Ait-Aoudia [a]

[a] École Nationale Supérieure en Informatique, Algiers, Algeria
[b] Université du Québec en Outaouais, Gatineau, QC, Canada

ARTICLE INFO

Article history:
Received 4 February 2013
Received in revised form
13 July 2013
Accepted 10 September 2013
Available online 19 September 2013

Keywords:
Image segmentation
Multi-modal class thresholding
Mixture of generalized Gaussian distributions (MoGG)

ABSTRACT

This paper presents a new approach to multi-class thresholding-based segmentation. It considerably improves existing thresholding methods by efficiently modeling non-Gaussian and multi-modal class-conditional distributions using mixtures of generalized Gaussian distributions (MoGG). The proposed approach seamlessly: (1) extends the standard Otsu's method to arbitrary numbers of thresholds and (2) extends the Kittler and Illingworth minimum error thresholding to non-Gaussian and multi-modal class-conditional data. MoGGs enable efficient representation of heavy-tailed data and multi-modal histograms with flat or sharply shaped peaks. Experiments on synthetic data and real-world image segmentation show the performance of the proposed approach with comparison to recent state-of-the-art techniques.

1. Introduction

Thresholding-based image segmentation is a well-known technique that is used in a broad range of applications, such as change detection [20], object recognition [3,34] and document image analysis [26], to name a few. Image thresholding aims at building a partition of an image into K classes, $C_1,\ldots,C_K$, which are separated by $K-1$ thresholds $T_1,\ldots,T_{K-1}$. In case of $K=2$, the image is segmented into foreground and background regions. In case of $K>2$, the image is segmented into K distinct regions. In most of existing thresholding methods, the parameter K is generally given and it corresponds to the number of histogram modes [27]. Comparative studies about existing thresholding techniques applied to image segmentation can be found in [10,24,27,32].

Among the most popular methods for image thresholding are the standard Otsu's method [22] and Kittler and Illingworth's method [14]. While the former uses inter-class separability to calculate optimal thresholds between classes, the latter is based on the minimization of Bayes classification error, where each class is modeled by a Gaussian distribution. Both methods, however, assume a uni-modal shape for classes and use sample mean and standard deviation (i.e., the parameters of a Gaussian) to approximate their distributions. In [12,32], the authors established the relationship between the two methods, where these parameters can be obtained in either methods using maximum likelihood estimation of a Gaussian model for each class. *Entropy* and *relative entropy* can also be used to derive good thresholds for image segmentation when the distribution of classes is Gaussian [5,6,25]. For example, Jiulun and Winxin [12] gave a relative-entropy interpretation for the minimum error thresholding (MET) [14,19]. In that work, the Kullback–Leibler divergence [15] is used to measure the discrepancy between histograms of a source image and a mixture of two Gaussians. Recently, Xue and Titterington [30] proposed a thresholding method where class data are modeled by Laplacian distributions. They showed that the obtained thresholds offer better separation of classes when their distributions are skewed, heavy-tailed or contaminated by outliers. Indeed, the location and dispersion parameters of the Laplacian distribution are the median and the absolute deviation from the median, which are more robust to outliers compared to the sample mean and standard deviation, respectively [11].

Previous methods for image thresholding were basically devised to separate classes that are unimodal [27]. Therefore, they are not adapted to multi-modal class segmentation. For instance, in many segmentation problems (e.g., medical images, and remote sensing), one might want to separate the image foreground from a background region, each of which may have a multi-modal distribution. Another limitation for the standard methods [14] and [22], and as pointed out in [30], lies in the assumption that class data are Gaussian. In several image examples, one can find histogram modes that are skewed, sharply peaked or heavy tailed, making the assumption of Gaussian-distributed classes not realistic. Recently, researchers have used other distribution types to

* Corresponding author. Tel.: +1 819 595 3900x1601.
E-mail addresses: a_boulmerka@esi.dz (A. Boulmerka), mohandsaid.allili@uqo.ca (M. Saïd Allili), s_ait_aoudia@esi.dz (S. Ait-Aoudia).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.09.004

provide better image thresholding methods by modeling histogram classes using, for instance, Poisson [23], generalized Gaussian [2,7,8], skew-normal [31] and Rayleigh [29] distributions. However, these approaches are also built on the assumption that all classes are unimodal. Worth mentioning is the parallel trend of using mixture methods for segmentation (ex. [1,21,35,36]), where data are clustered to classes determined by the components of a learned mixture model. For such works, the number of classes (which correspond to the number of mixture components) can be estimated using information-theoretic criteria such as AIC, BIC, MML, etc. [18]. This paper deals with a different problem which consists of finding thresholds between classes with distributions that can be constituted of arbitrary numbers of (non-Gaussian) histogram modes. Thus, contrary to [1,21], the number of classes K will not necessarily correspond to the number histogram modes.

In this paper, we propose a new thresholding approach that performs segmentation for multi-modal classes with arbitrarily shaped modes. We generalize the aforementioned state-of-art techniques, based on using single probability density functions (pdf's), to mixtures of generalized Gaussian distributions (MoGG's). The Generalized Gaussian Distributions (GGD) is a generalization of the Laplacian and the normal distributions in that it has an additional degree of freedom that controls its kurtosis. Therefore, histogram modes, ranging from sharply peaked to flat ones, can be accurately represented using this model. Furthermore, skewed and multi-modal classes are accurately represented using mixtures of GGDs. We propose an objective function that finds optimal thresholds for multi-modal classes of data. It also extends easily to arbitrary numbers of classes ($K>2$) with reasonable computational time. Experiments on synthetic data, as well as real-world image segmentation, show the performance of the proposed approach.

This paper is organized as follows: Section 2 presents state-of-the-art theory for thresholding techniques. In Section 3 we outline our proposed approach for image thresholding. Experimental results are given in Section 4. We end the paper with a conclusion and some future work perspectives.

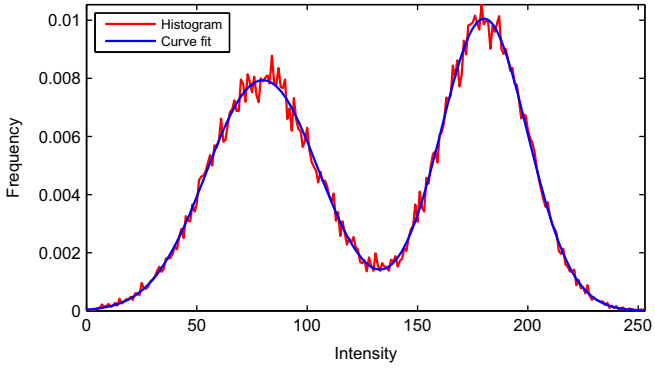

Fig. 1. Bimodal histogram ($K=2$).

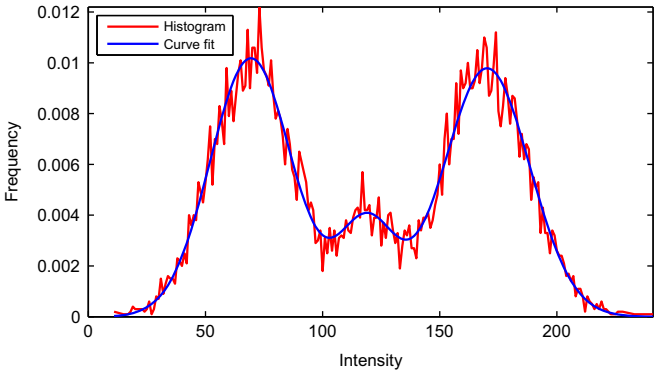

Fig. 2. Multimodal histogram ($K=3$).

2. General formulation of the Otsu's method (case K=2)

Let $\mathcal{X}=\{x_1,x_2,\ldots,x_N\}$ be the gray levels of the pixels of an image I of size $N=H\times W$; H and W being the height and the width of the image. Let $\mathbf{t}=(t_1,t_2,\ldots,t_{K-1})$ be a set of thresholds that partitions an image into K classes. First we consider the simple case of $K=2$. The most general case of $K>2$ will be elaborated later in this paper. In the case of $K=2$, one threshold t yields two classes $\mathcal{C}_1(t)=\{x:0\le x\le t\}$ and $\mathcal{C}_2(t)=\{x:t+1\le x\le T\}$, where T is the maximum gray level. Finally, we denote by $h(x)$ the histogram frequency of the gray level x, where $\sum_{x=0}^{T}h(x)=1$. The resulting histogram in this case ($K=2$) is bimodal, as shown in Fig. 1. Otsu's method [22] determines the optimal threshold t using discriminant analysis, by maximizing inter-class variation, or equivalently minimizing intra-class variation.

A generalized formula of the Otsu's method for $K=2$ can be defined as follows (see refs. [10,14,22,27,30,32]):

$$\sigma_B^2(t)=\arg\min_t\{\omega_1(t)V_1(t)+\omega_2(t)V_2(t)\}, \tag{1}$$

where, we have

$$\begin{cases}\omega_1(t)=\sum_{x=0}^{t}h(x)\\ \omega_2(t)=\sum_{x=t+1}^{T}h(x)=1-\omega_1(t)\end{cases}, \tag{2}$$

and

$$\begin{cases}V_1(t)=\dfrac{1}{\omega_1(t)}\sum_{x=0}^{t}h(x)\|(x-m_1(t)\|_\beta\\ V_2(t)=\dfrac{1}{\omega_2(t)}\sum_{x=t+1}^{T}h(x)\|(x-m_2(t)\|_\beta\end{cases}, \tag{3}$$

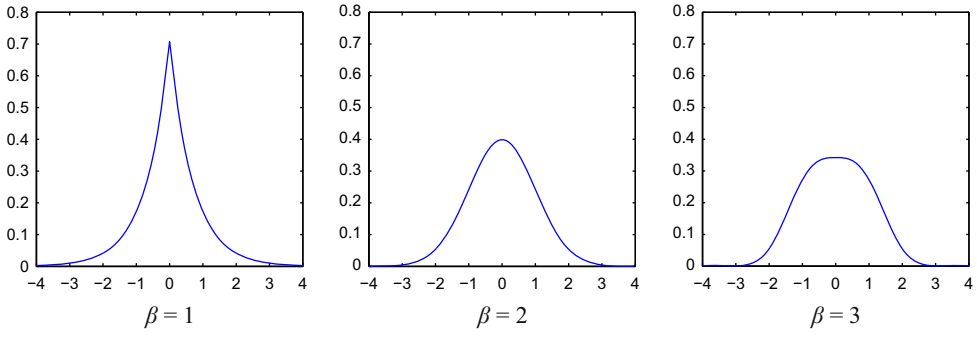

Fig. 3. Different shapes of the GGD distribution as a function of the parameter β (μ=0,σ=1).

where $\|\cdot\|$ is the norm symbol. When $\beta=2$, the model corresponds to the standard Otsu's method and the minimum error thresholding proposed in [14] and [22], respectively. When $\beta=1$, the model corresponds to the method proposed in [30]. In the case $\beta=2$, the estimated location parameters $m_1(t)$ and $m_2(t)$ correspond to the sample means of the classes $\mathcal{C}_1$ and $\mathcal{C}_2$; whereas, in the second case $\beta=1$, these parameters correspond to the sample medians of the classes, as proposed in [30,32]. For multi-thresholding, since the classical Ostu's method, extensions were proposed to arbitrary number of classes (see the next section). However, all these works assume that the classes follow unimodal distributions with their data generally represented by their mean or median parameters.

2.1. Standard Otsu's method (case $\beta=2$)

This method, proposed in [22], consists of calculating an optimal threshold t that segments an image into two distinct regions using the following minimization:

$$t=\arg\min_{t}\{\omega_1(t)s_1^2(t)+\omega_2(t)s_2^2(t)\}, \tag{4}$$

where $\omega_1(t)$ and $\omega_2(t)$, defined in Eq. (2), correspond to proportions of pixels representing classes $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively. The parameters $s_1(t)$ and $s_2(t)$ represent sample standard deviations for the classes $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively, and are defined as follows:

$$\begin{cases} s_1^2(t)=\sum_{x=0}^{t}\left[\frac{h(x)}{\omega_1(t)}(x-\bar{x}_1(t))^2\right] \\ s_2^2(t)=\sum_{x=t+1}^{T}\left[\frac{h(x)}{\omega_2(t)}(x-\bar{x}_2(t))^2\right] \end{cases}, \tag{5}$$

where $\bar{x}_1(t)$ and $\bar{x}_2(t)$ correspond to the sample means for $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively, which are defined as $\bar{x}_1(t)=\sum_{x=0}^{t}(h(x)/\omega_1(t))x$ and $\bar{x}_2(t)=\sum_{x=t+1}^{T}(h(x)/\omega_2(t))x$.

It can be shown that the Otsu's method formulation can be derived from maximization of the log-likelihood of a mixture of two Gaussian distributions [16]. Finally, this method can be easily extended to multi-level thresholding. Given that the image histogram contains K unimodal classes (see Fig. 2), then $K-1$ thresholds are required to separate the classes $\mathcal{C}_1$, $\mathcal{C}_2$,…, $\mathcal{C}_K$ that correspond to gray level intervals $[0,t_1]$, $[t_{1+1},t_2]$, …, $[t_{K-1},T]$, respectively. More formally, multi-level thresholds $\mathbf{t}=(t_1,\ldots,t_{K-1})$ are obtained using the following minimization [14,22,27]:

$$\mathbf{t}=\arg\min_{\mathbf{t}}\sum_{k=1}^{K}\{\omega_k(\mathbf{t})s_k^2(\mathbf{t})\}, \tag{6}$$

where $s_k^2(\mathbf{t})=\sum_{x=t_{k-1}}^{t_k}[h(x)/\omega_k(t)(x-\bar{x}_k(\mathbf{t}))^2]$ and $\bar{x}_k(\mathbf{t})=\sum_{x=t_{k-1}}^{t_k}(h(x)/\omega_k(\mathbf{t}))x$. Note that Kittler and Illingworth's method [14] uses a slightly modified formula, where an optimal threshold vector $\mathbf{t}$ is selected for arbitrary K, $K\geq 2$, as follows:

$$\mathbf{t}=\arg\min_{\mathbf{t}}\sum_{k=1}^{K}\omega_k(\mathbf{t})\log\frac{s_k^2(\mathbf{t})}{\omega_k(\mathbf{t})}, \tag{7}$$

Table 1
Misclassification Errors obtained for the NDT-images. Columns from left to right show: the standard Otsu's, the median extension, the MoG and the MoGG methods, respectively. Rows from top to bottom show, respectively, the average, the standard deviation, the minimum and the maximum values of ME obtained by each method.

	Classical modeling		Mixture modeling	
	Standard Otsu's	Median extension	MoG	MoGG
AVG	0.1767	0.2315	0.0149	0.0115
STD	0.2103	0.2389	0.0137	0.0106
MIN	0.0003	0.0003	0.0001	0.0001
MAX	0.6261	0.6085	0.0486	0.0381

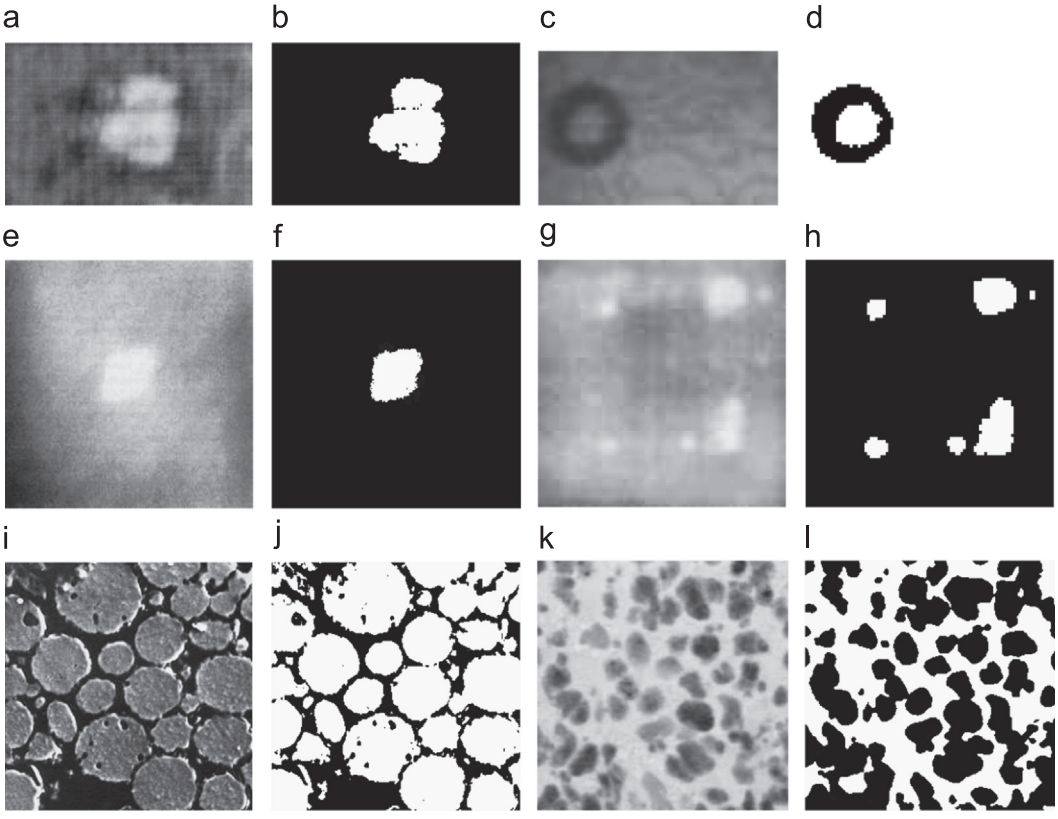

Fig. 4. A sample of NDT-images and their ground truth segmentation: (a,b) ultrasonic GFRP material, (c,d) Eddy current image, (e,f) thermal image of GFRP composite material, (g,h) defective eddy current image, (i,j) light microscopic image of a material structure and (k,l) material structure image.

2.2. Median-based extension of Otsu's method (case $\beta=1$)

Recently, Xue and Titterington [30] proposed to use the median parameter instead of the mean in the Otsu's method formulation. For the general case of multi-level thresholding, the optimal threshold $\mathbf{t}=(t_1,\ldots,t_{K-1})$ can be estimated through the following rule:

$$\mathbf{t}=\arg\min_{\mathbf{t}}\sum_{k=1}^{K}\{\omega_k(\mathbf{t})MAD_k(\mathbf{t})\}, \tag{8}$$

where $MAD_k(t)$ represents the absolute deviation from the median for class $C_k(\mathbf{t})$ $(k=1,\ldots,K)$, which is given by $MAD_k(\mathbf{t})=\sum_{x=t_{k-1}}^{t_k} h(x)/\omega_k(\mathbf{t})|(x-m_k(\mathbf{t})|$, where $m_k(\mathbf{t})=\text{median}\{x\in C_k(\mathbf{t})\}$. Using this formulation, the authors proved that the median-based approach can be derived by modeling each class using the Laplacian distribution. They also showed the robustness of using the median for obtaining optimal thresholds when class-conditional distributions are skewed or contaminated by outliers. However, when one of the classes contains multiple modes, the models underlying Eq. (1), and therefore Eq. (8), will not be applicable since it is based on intra-class variation which is built on the assumption that classes are unimodal.

3. Multi-modal class thresholding

We propose to extend the general model formulated by Eqs. (1)–(3) to represent multi-modal class-conditional distributions using finite mixture models (FMM). Finite mixtures are a flexible and powerful probabilistic tool for modeling univariate

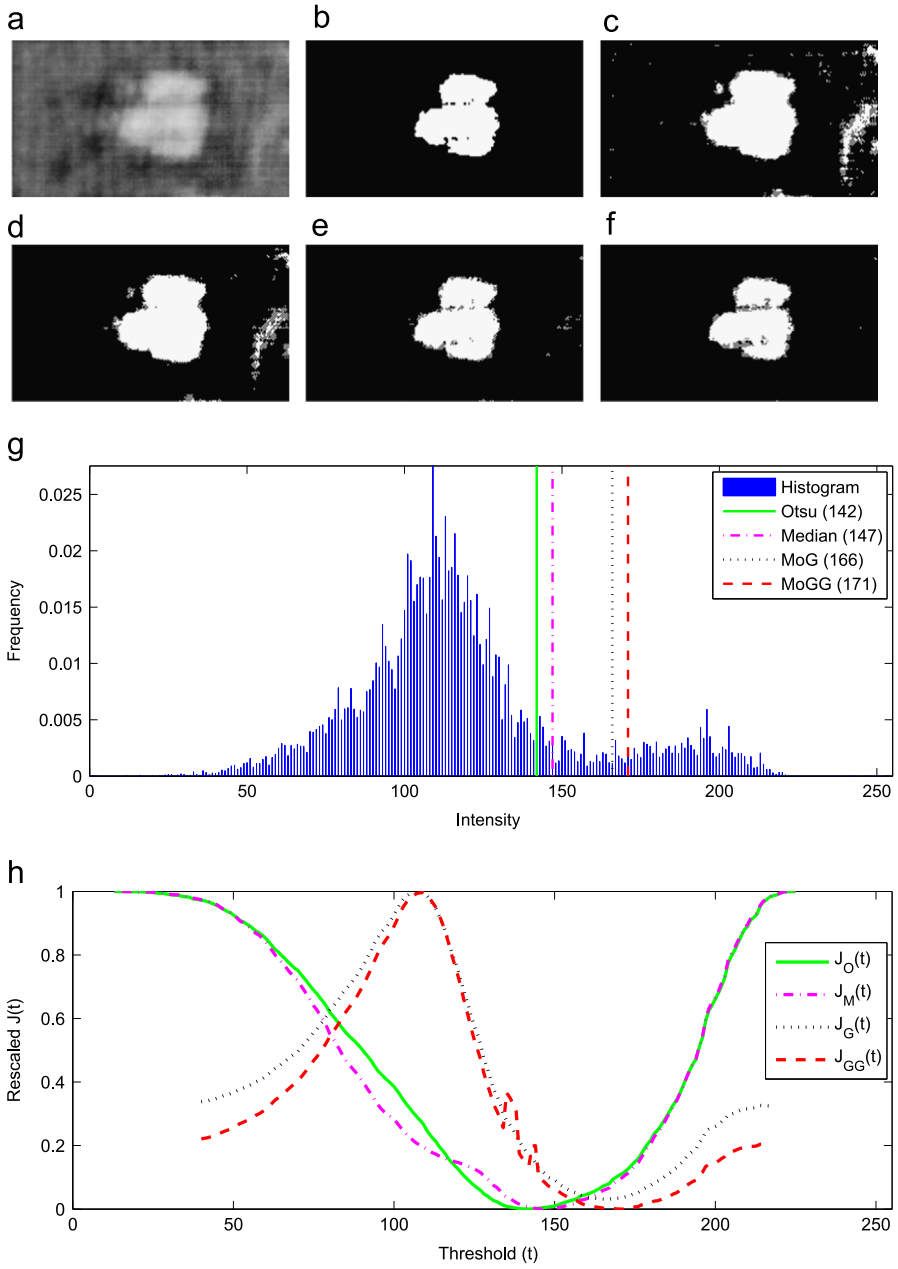

Fig. 5. For 'NDT-image1' we present (a) the original image, (b) the ground truth, (c)–(f) segmentation results obtained by the standard Otsu's, the median extension, MoG and MoGG methods, respectively, (g) image histogram superimposed with optimal thresholds ($\mathbf{t}^*_O$, $\mathbf{t}^*_M$, $\mathbf{t}^*_G$ and $\mathbf{t}^*_{GG}$) and (h) plots of functions $J^*_O(\mathbf{t})$, $J^*_M(\mathbf{t})$, $J^*_G(\mathbf{t})$ and $J^*_{GG}(\mathbf{t})$ corresponding to optimal thresholds in the above histogram.

and multivariate data [9,18]. They allow for modeling randomly generated data from multiple sources in an unsupervised fashion. Recently, Allili et al. [1] proposed to use finite mixtures of generalized Gaussian distributions (MoGG) to model non-Gaussian data containing multiple classes. Indeed, the generalized Gaussian density (GGD) is an extension of the Gaussian and the Laplacian distributions [4] and has the following formulation:

$$p(x|\mu,\sigma,\beta) = K(\beta,\sigma)\exp(-A(\beta)|(x-\mu)/\sigma|^{\beta}) \tag{9}$$

where $K(\beta,\sigma) = \beta\sqrt{\Gamma(3/\beta)/\Gamma(1/\beta)}/(2\sigma\Gamma(1/\beta))$ and $A(\beta) = [\Gamma(3/\beta)/\Gamma(1/\beta)]^{\beta/2}$; $\Gamma(.)$ being the gamma function. The parameters μ and σ are the GGD location and dispersion parameters. The parameter β controls the kurtosis of the pdf and determines whether it is peaked or flat: the larger the value of β, the flatter the pdf; and the smaller β is, the more peaked the pdf. This gives the pdf a flexibility to fit the shape of heavy-tailed data [4]. Two well-known special cases of the GGD model are the Laplacian, as $\beta=1$, and the Gaussian distribution, as $\beta=2$ (see Fig. 3).

3.1. Multimodal class thresholding (case K=2)

Here, we suppose that the data consist of two classes $\mathcal{C}_1$ and $\mathcal{C}_2$ separated by a candidate threshold t, each of which is multi-modal and modeled using a MoGG, as follows:

$$\begin{cases} p(x|\mathcal{C}_1) = \sum_{k=1}^{K_1} \alpha_k(t)\, p(x|\vec{\theta}_k(t)), & (10) \\ p(x|\mathcal{C}_2) = \sum_{j=1}^{K_2} \pi_j(t)\, p(x|\vec{\varphi}_j(t)), & (11) \end{cases}$$

where $\alpha_k(t)$, $k\in\{1,\ldots,K_1\}$ and $\pi_j(t)$, $j\in\{1,\ldots,K_2\}$, are the mixing parameters of the two mixtures, such that $\sum_{k=1}^{K_1}\alpha_k(t)=1$ and

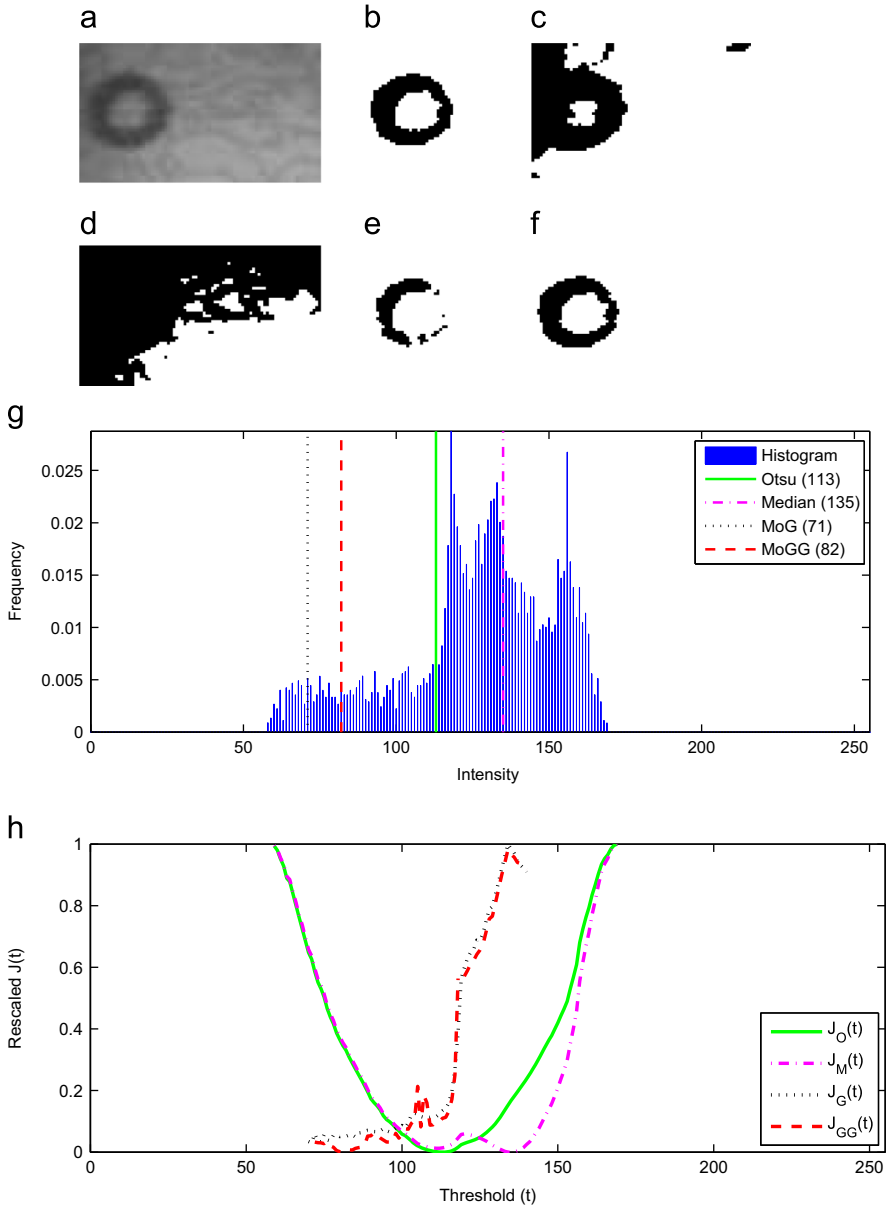

Fig. 6. For 'NDT-image8'. Caption is as for Fig. 5.

$\sum_{j=1}^{K_2}\pi_j(t)=1$. The number of components K_1 and K_2 are application specific (e.g., segmentation of the brain cortex in medical images, foreground/background segmentation, etc). They can be empirically estimated using a sample of data (e.g., a set of images of the same type). Each component in Eqs. (10) and (11), $p(x|\vec{\theta}_k(t))$ and $p(x|\vec{\varphi}_j(t))$ is a GGD. Using our mixture modelling on the image data $\mathcal{X}$, Eq. (1) can be expressed as a maximization of class-likelihoods, as follows:

$$t=\arg\max_t\{\omega_1(t)L_1(t)+\omega_2(t)L_2(t)\}, \tag{12}$$

Table 2
Parameters setting for bimodal simulated dataset benchmarks.

Bench.	Class	K_r	μ	σ	β	π
Bench.1	$r=1$	2	$\mu_{1,1}=50$	$\sigma_{1,1}=20$	$\beta_{1,1}=1.00$	$\pi_{1,1}=0.40$
			$\mu_{1,2}=100$	$\sigma_{1,2}=20$	$\beta_{1,2}=2.00$	$\pi_{1,2}=0.40$
	$r=2$	2	$\mu_{2,1}=150$	$\sigma_{2,1}=20$	$\beta_{2,1}=4.00$	$\pi_{2,1}=0.10$
			$\mu_{2,2}=150$	$\sigma_{2,2}=20$	$\beta_{2,2}=4.00$	$\pi_{2,2}=0.10$
Bench.2	$r=1$	2	$\mu_{1,1}=70$	$\sigma_{1,1}=20$	$\beta_{1,1}=1.00$	$\pi_{1,1}=0.10$
			$\mu_{1,2}=70$	$\sigma_{1,2}=20$	$\beta_{1,2}=2.00$	$\pi_{1,2}=0.10$
	$r=2$	2	$\mu_{2,1}=120$	$\sigma_{2,1}=20$	$\beta_{2,1}=1.00$	$\pi_{2,1}=0.20$
			$\mu_{2,2}=170$	$\sigma_{2,2}=20$	$\beta_{2,2}=4.00$	$\pi_{2,2}=0.60$
Bench.3	$r=1$	2	$\mu_{1,1}=50$	$\sigma_{1,1}=20$	$\beta_{1,1}=1.00$	$\pi_{1,1}=0.10$
			$\mu_{1,2}=50$	$\sigma_{1,2}=20$	$\beta_{1,2}=2.00$	$\pi_{1,2}=0.10$
	$r=2$	2	$\mu_{2,1}=100$	$\sigma_{2,1}=20$	$\beta_{2,1}=2.00$	$\pi_{2,1}=0.20$
			$\mu_{2,2}=160$	$\sigma_{2,2}=20$	$\beta_{2,2}=4.00$	$\pi_{2,2}=0.60$
Bench.4	$r=1$	2	$\mu_{1,1}=100$	$\sigma_{1,1}=20$	$\beta_{1,1}=1.00$	$\pi_{1,1}=0.20$
			$\mu_{1,2}=100$	$\sigma_{1,2}=20$	$\beta_{1,2}=4.00$	$\pi_{1,2}=0.10$
	$r=2$	2	$\mu_{2,1}=150$	$\sigma_{2,1}=30$	$\beta_{2,1}=4.00$	$\pi_{2,1}=0.30$
			$\mu_{2,2}=180$	$\sigma_{2,2}=20$	$\beta_{2,2}=1.00$	$\pi_{2,2}=0.40$
Bench.5	$r=1$	2	$\mu_{1,1}=60$	$\sigma_{1,1}=20$	$\beta_{1,1}=2.00$	$\pi_{1,1}=0.05$
			$\mu_{1,2}=70$	$\sigma_{1,2}=10$	$\beta_{1,2}=1.00$	$\pi_{1,2}=0.05$
	$r=2$	3	$\mu_{2,1}=100$	$\sigma_{2,1}=30$	$\beta_{2,1}=4.00$	$\pi_{2,1}=0.20$
			$\mu_{2,2}=150$	$\sigma_{2,2}=20$	$\beta_{2,2}=2.00$	$\pi_{2,2}=0.35$
			$\mu_{2,3}=170$	$\sigma_{2,3}=20$	$\beta_{2,3}=1.00$	$\pi_{2,3}=0.35$

where:

$$\begin{cases} L_1(t)=\prod_{0\le x_i\le t}\left\{\sum_{k=1}^{K_1}\alpha_k(t)p(x_i|\vec{\theta}_k(t))\right\}, & (13)\\ L_2(t)=\prod_{t+1\le x_i\le T}\left\{\sum_{j=1}^{K_2}\pi_j(t)p(x_i|\vec{\varphi}_j(t))\right\}. & (14)\end{cases}$$

By taking the logarithm of each term in Eq. (12) and writing the results in term of gray level frequencies, we obtain

$$\begin{cases} \log(\omega_1(t)L_1(t))=N\sum_{x=0}^{t}h(x)\log\left\{\omega_1(t)\sum_{k=1}^{K_1}\alpha_k(t)p(x|\vec{\theta}_k(t))\right\}, & (15)\\ \log(\omega_2(t)L_2(t))=N\sum_{x=t+1}^{T}h(x)\log\left\{\omega_2(t)\sum_{j=1}^{K_2}\pi_j(t)p(x|\vec{\varphi}_j(t))\right\} & (16).\end{cases}$$

Finally, the parameters $(\alpha_k(t),\vec{\theta}_k(t))$, $k\in\{1,\ldots,K_1\}$, and $(\pi_j(t),\vec{\varphi}_j(t))$, $j\in\{1,\ldots,K_2\}$, are estimated using the maximum likelihood method. We use the Expectation–Maximization (EM) algorithm to obtain the parameters of the mixture of generalized Gaussian distributions, as proposed in [1] (we refer the reader to that reference for the derivation of the E–M steps).

3.2. Multimodal class thresholding (Case $K>2$)

When there are more than two classes in the image, one can readily generalize the two-class case model of Eq. (12) using a

Table 3
Classification error in the simulated data sets (between brackets is the percentage of data sets where MoGG gives the least error).

Bench.	Error otsu	Error median	Error MoG	Error MoGG
1	0.188 (100%)	0.182 (100%)	0.126 (97%)	0.093
2	0.187 (100%)	0.199 (100%)	0.086 (89%)	0.035
3	0.168 (99%)	0.173 (100%)	0.152 (85%)	0.062
4	0.168 (67%)	0.159 (63%)	0.190 (80%)	0.146
5	0.181 (100%)	0.232 (100%)	0.124 (92%)	0.089

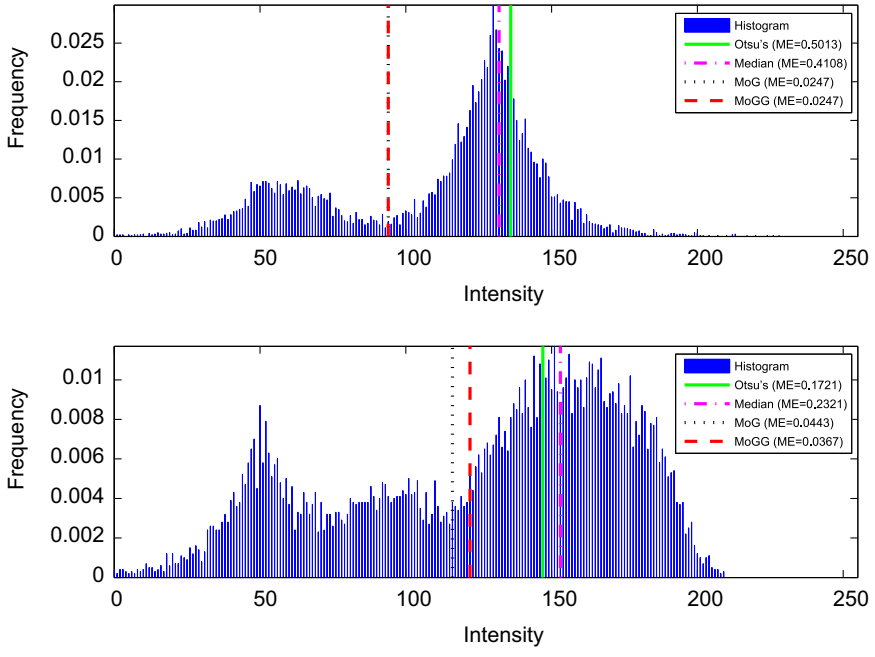

Fig. 7. Examples of generated data sets histograms with optimal thresholding results and misclassification errors (ME) given by different methods (original Otsu's method, median based extension method, MoG method and MoGG method.).

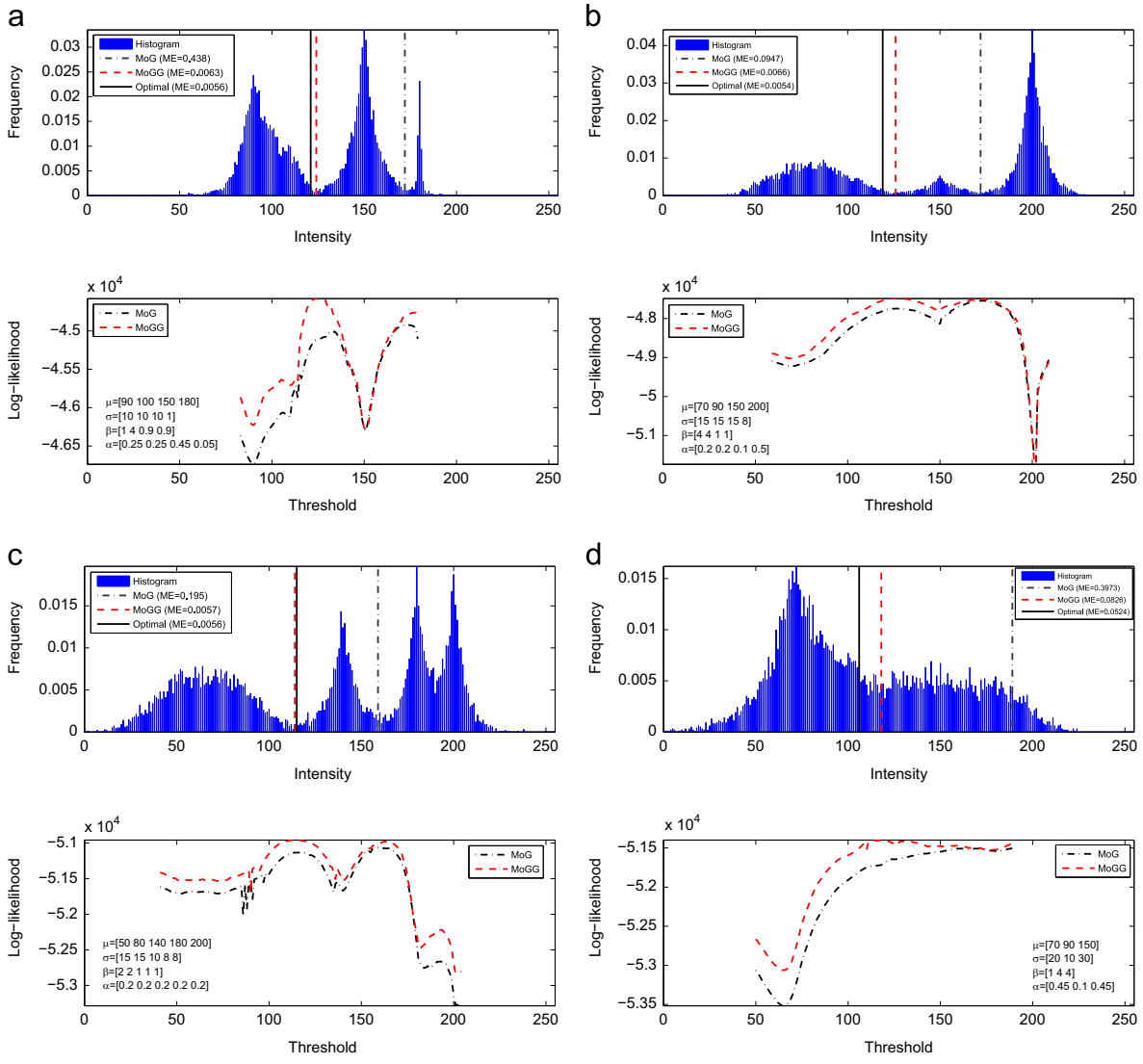

Fig. 8. Optimal thresholds, misclassification error and log-likelihoods obtained using MoG and MoGG methods for 4 synthetic generalized Gaussian mixture densities inspired from Gaussian mixture densities given by [18] (Chapter 1, page 12). (a) $K_{left}=2$, $K_{right}=2$, (b) $K_{left}=2$, $K_{right}=2$, (c) $K_{left}=2$, $K_{right}=3$ and (d) $K_{left}=2$, $K_{right}=1$.

Table 4
Parameters setting for bimodal simulated dataset benchmarks (MoG Versus MoGG experimentation).

Bench.	Class	K_r	μ	σ	β	π
Bench.1	$r=1$	2	$\mu_{1,1}=70$	$\sigma_{1,1}=15$	$\beta_{1,1}=4.00$	$\pi_{1,1}=0.15$
			$\mu_{1,2}=90$	$\sigma_{1,2}=15$	$\beta_{1,2}=4.00$	$\pi_{1,2}=0.15$
	$r=2$	2	$\mu_{2,1}=120$	$\sigma_{2,1}=15$	$\beta_{2,1}=1.00$	$\pi_{2,1}=0.30$
			$\mu_{2,2}=140$	$\sigma_{2,2}=8$	$\beta_{2,2}=1.00$	$\pi_{2,2}=0.40$
Bench.2	$r=1$	2	$\mu_{1,1}=70$	$\sigma_{1,1}=25$	$\beta_{1,1}=4.00$	$\pi_{1,1}=0.30$
			$\mu_{1,2}=90$	$\sigma_{1,2}=15$	$\beta_{1,2}=4.00$	$\pi_{1,2}=0.10$
	$r=2$	2	$\mu_{2,1}=120$	$\sigma_{2,1}=15$	$\beta_{2,1}=1.00$	$\pi_{2,1}=0.20$
			$\mu_{2,2}=130$	$\sigma_{2,2}=8$	$\beta_{2,2}=1.00$	$\pi_{2,2}=0.40$
Bench.3	$r=1$	2	$\mu_{1,1}=80$	$\sigma_{1,1}=10$	$\beta_{1,1}=4.00$	$\pi_{1,1}=0.25$
			$\mu_{1,2}=100$	$\sigma_{1,2}=10$	$\beta_{1,2}=4.00$	$\pi_{1,2}=0.25$
	$r=2$	2	$\mu_{2,1}=150$	$\sigma_{2,1}=15$	$\beta_{2,1}=1.00$	$\pi_{2,1}=0.25$
			$\mu_{2,2}=200$	$\sigma_{2,2}=8$	$\beta_{2,2}=1.00$	$\pi_{2,2}=0.25$

vector of thresholds $\mathbf{t}=\{t_1,\ldots,t_{K-1}\}$. Therefore, the maximization in Eq. (12) will be reformulated as follows:

$$\mathbf{t}=\arg\max_{\mathbf{t}}\left\{\sum_{r=1}^{K}\omega_r(\mathbf{t})L_r(\mathbf{t})\right\}, \tag{17}$$

Table 5
Classification error in the simulated data sets of Table 4 benchmarks (between brackets is the percentage of data sets where MoGG gives the least error).

Bench.	Error MoG	Error MoGG
1	0.231 (85%)	0.081
2	0.344 (100%)	0.069
3	0.240 (100%)	0.010

where, we have

$$L_r(\mathbf{t})=\prod_{a_r\le x_i\le b_r}\left\{\sum_{k=1}^{K_r}\alpha_{r,k}(\mathbf{t})p(x_i|\vec{\theta}_{r,k}(\mathbf{t}))\right\}, \tag{18}$$

and

$$a_r=\begin{cases}0 & \text{if } r=1\\ t_{r-1} & \text{if } r>1\end{cases} \tag{19}$$

$$b_r=\begin{cases}T & \text{if } r=K-1\\ t_r & \text{if } r<K-1\end{cases} \tag{20}$$

After taking the logarithm of each class-likelihood and writing the result in terms of gray level frequencies, we obtain

$$\log(\omega_r(\mathbf{t})L_r(\mathbf{t})) = N\sum_{x=a_r}^{b_r} h(x)\log\left\{\omega_r(\mathbf{t})\sum_{k=1}^{K_r}\alpha_{r,k}(\mathbf{t})p(x|\vec{\theta}_{r,k}(\mathbf{t}))\right\}. \quad (21)$$

The advantage of using Eq. (17) is that it can segment classes with multi-modal distributions, which is impossible to achieve using the thresholding formulation given by Eq. (1). In fact, Eq. (1) can be obtained as a special case of Eq. (17) where each class is modeled as a mixture of a single GGD component with unit variance. Note that this advantage comes with additional time cost to estimate a mixture model parameters for each class using the maximum likelihood method (MLE). Nonetheless, this additional time is not a limitation for our approach given the high thresholding accuracy and flexibility it offers.

The procedure for estimating the optimal threshold vector $\mathbf{t}$ calculates iteratively for each threshold t_r ($r=1,\ldots,K-1$). For each threshold t_r, we consider a left class $\mathcal{C}^r_{left}$ and a right class $\mathcal{C}^r_{right}$, represented each by a mixture of K^r_{left} and K^r_{right} GGDs, respectively. The maximum likelihood estimation is then performed for the parameters of $\mathcal{C}^r_{left}$ and $\mathcal{C}^r_{right}$ using the Expectation-Maximization (EM) algorithm [1]. Note that to speed up the EM estimation for a candidate threshold t, we use the estimated parameters obtained for $t-1$ as initialization. Thus, a small number of iterations is required to adjust the MoGG parameters for each tested threshold. The different steps of the method for estimating the elements of $\mathbf{t}$ are explained in the script of Algorithm 1.

Algorithm 1. Compute $\mathbf{t}=\{t_1,\ldots,t_{K-1}\}$ using Eq. (17).

Inputs:
- Image histogram,
- Number of classes K,
- Number of components of each class $\{K_i, i=1,\ldots K\}$.

Output:
- Threshold vector $\mathbf{t}=\{t_1,\ldots,t_{K-1}\}$.

$r \leftarrow 1$; // *the first threshold*
while $r \le K-1$ **do**
  $K^r_{left} \leftarrow \sum_{i=1}^{r} K_i$;
  $K^r_{right} \leftarrow \sum_{j=r+1}^{K} K_j$;
  if$(r=1)$ **then**
    $s_1 \leftarrow 1$; //*minimum gray level* $+1$

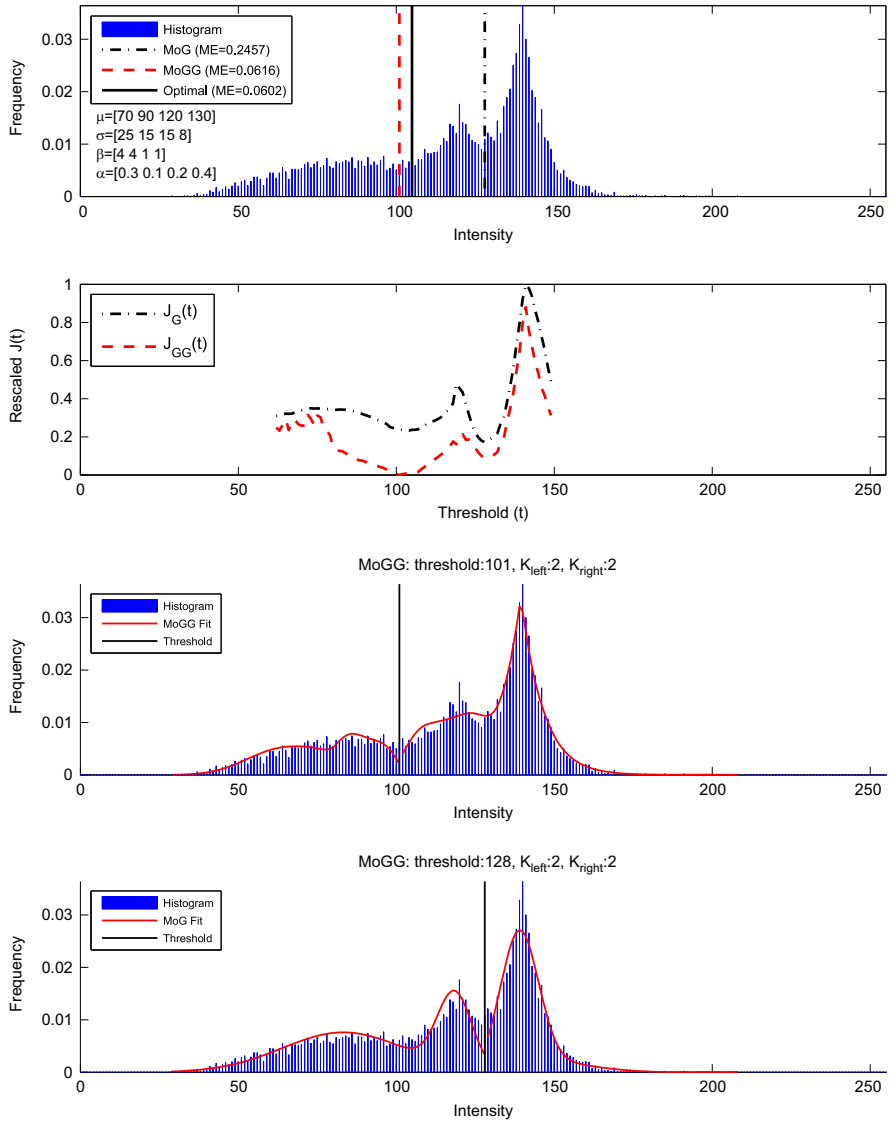

Fig. 9. Example of a generated data set histogram from the benchmark 1 of Table 5 with optimal thresholding results and misclassification errors (ME) given by MoG and MoGG methods.

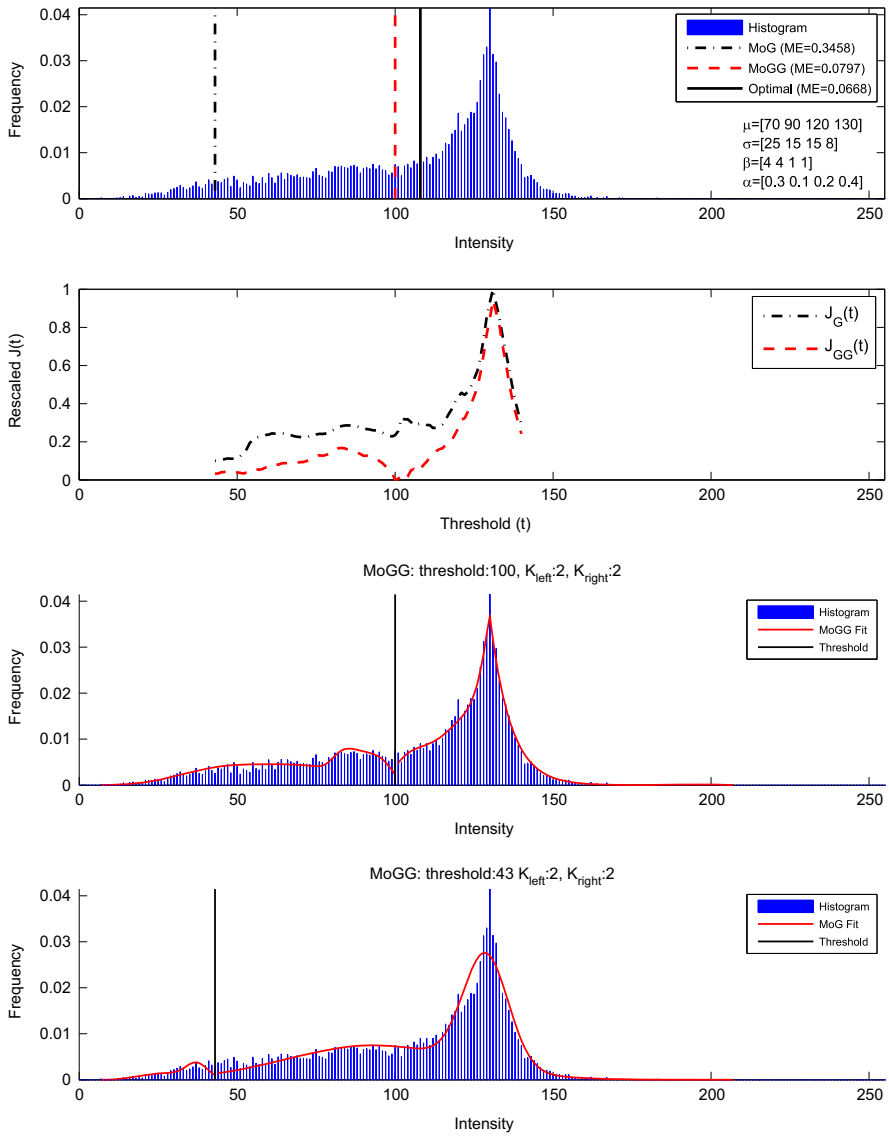

Fig. 10. Example of a generated data set histogram from the benchmark 2 of Table 5 with optimal thresholding results and misclassification errors (ME) given by MoG and MoGG methods.

else
$s_1 \leftarrow t_{r-1}$;
end if
$s_2 \leftarrow T-1$; //*maximum gray level*-1
for $t = s_1 \rightarrow s_2$ **do**
- Calculate $\omega_1(t)$ and $\omega_2(t)$;
- Estimate the parameters of a mixture of K_{left} GGDs (class $\mathcal{C}_{left}$);
- Estimate the parameters of a mixture of K_{right} GGDs (class $\mathcal{C}_{right}$);
- Calculate $J_r(t) = \log(\omega_1(t)L_1(t)) + \log(\omega_2(t)L_2(t))$;
end for
$t_r \leftarrow \text{argmin}_t J_r(t)$;
$r \leftarrow r+1$; // *the next threshold*
end while

4. Experimental results

We conducted several experiments to measure the performance of the proposed approach by comparing it to recent state-of-the-art thresholding methods. For this purpose, we used synthetic histograms as well as real images from known datasets [27]. Quantitative results are presented showing how well the proposed model finds optimal thresholds in terms of segmentation accuracy.

We objectively measure thresholding performance by using Misclassification Error (ME) criterion. For foreground/background image segmentation, the ME reflects the percentage of misclassified pixels, expressed as follows:

$$ME = 1 - \frac{|B_O \cap B_T| + |F_O \cap F_T|}{|B_O| + |F_O|}, \tag{22}$$

where B_O and F_O denote, respectively, the background and foreground of the original ground-truth image. B_T and F_T denote, respectively, the background and foreground pixels in the segmented image, where $|.|$ denotes set cardinality. The ME varies from 0, for a perfectly segmented image, to 1, for a totally wrongly segmented image.

The compared methods include (i) The standard Otsu's method [22], (ii) The median-based Otsu's extension method [30], (iii) Thresholding based on MoGs (Mixture of Gaussians) and (iv) Our thresholding method based on MoGGs. Note that the difference between MoG and MoGG methods is that for the MoGG method,

the shape parameter β is estimated using the EM algorithm (see Algorithm 1), while for the MoG method, the shape parameter is fixed to the Gaussian distribution (i.e., $\beta=2$). We compute the ME for all compared methods and rewrite for each method Eqs. (6), (8) and (17) as follows:

$$\mathbf{t}_O^* = \arg\min_{\mathbf{t}} J_O(\mathbf{t}), \tag{23}$$

$$\mathbf{t}_M^* = \arg\min_{\mathbf{t}} J_M(\mathbf{t}), \tag{24}$$

$$\mathbf{t}_G^* = \arg\min_{\mathbf{t}} J_G(\mathbf{t}), \tag{25}$$

$$\mathbf{t}_{GG}^* = \arg\min_{\mathbf{t}} J_{GG}(\mathbf{t}), \tag{26}$$

where $J_O(\mathbf{t})$ and $J_M(\mathbf{t})$ (standard Otsus's and median-based methods) are the terms to be minimized on the right-hand sides of Eqs. (6) and (8), $J_G(\mathbf{t})$ and $J_{GG}(\mathbf{t})$ (MoG and MoGG methods) are the terms to be maximized on the right-hand side of Eq. (17). To have better visualization for performance comparison, all values of the functions $J_O(\mathbf{t})$, $J_M(\mathbf{t})$, $J_G(\mathbf{t})$ and $J_{GG}(\mathbf{t})$ were re-scaled to the range $[0, 1]$.

4.1. Real-world image segmentation

We used the NDT-image dataset (NDT: Non Destructive Testing images) [27], which has also been used to evaluate some popular thresholding methods (e.g., [25,28,30]). The main properties of this dataset is that, on the one hand, it includes a variety of real-life thresholding applications like document image analysis, quality inspection of materials, defect detection, cell images, eddy current images, etc. On the other hand, it contains for each NDT-image the corresponding ground truth that considerably facilitates the evaluation task. Fig. 4 shows a sample of 6 images from this set that contains altogether 25 images.

Tests conducted on the NDT image dataset show that both MoGG and MoG methods give better results in terms of ME values against the standard Otsu's and the Median extension methods. Table 1 shows results obtained using the compared methods. We can note that on average, both MoG and MoGG have a lesser ME than the standard Otsu's and the Median extension methods. This performance of mixture-based methods is justified by their flexibility to represent multi-modal classes better than using class mean or median parameters used in the classical methods. Finally, we can note that thresholds given by the MoGG method are better

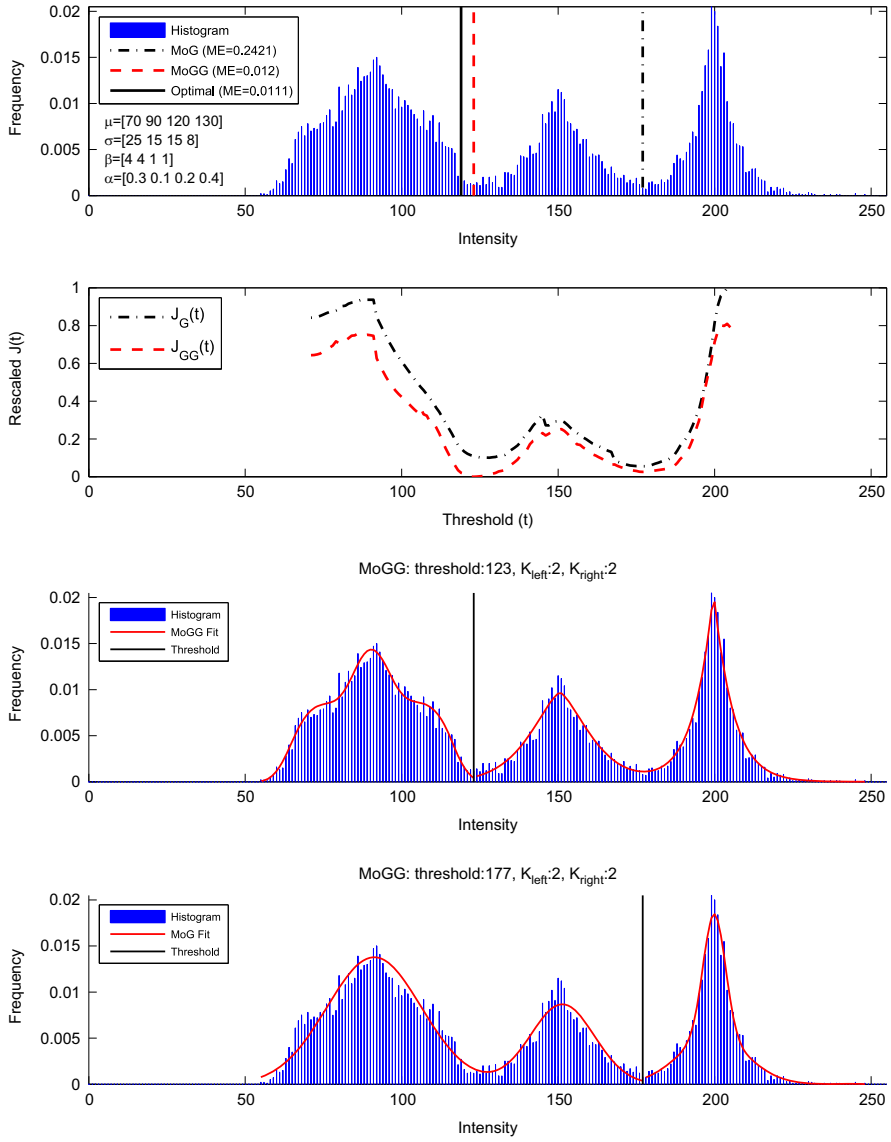

Fig. 11. Example of a generated data set histogram from the benchmark 3 of Table 5 with optimal thresholding results and misclassification errors (ME) given by MoG and MoGG methods.

than those obtained by the MoG method. This can be seen especially in histograms containing non-Gaussian modes, as will be illustrated in the following examples.

Figs. 5 and 6 show two examples of image segmentation (NDT-1 and NDT-8) using the compared methods. For each example, we show the segmentation ground truth, the image histogram with thresholds $\mathbf{t}_O^*, \mathbf{t}_M^*, \mathbf{t}_G^*$ and $\mathbf{t}_{GG}^*$, the plots for $J_O(\mathbf{t}), J_M(\mathbf{t}), J_G(\mathbf{t})$ and $J_{GG}(\mathbf{t})$, and the binarized image. For each example, the histogram should be separated into two classes (foreground and background). In the first example, the classes are approximately Gaussian, and all compared methods gave reasonable thresholds. However, using mixture models gave slightly better segmentations than using standard Otsu's and median extension methods. Finally, using MoGG gave slightly better segmentation than using MoG. In the second example, the classes are multi-modal and the mode shapes range from sharply peaked to flat ones. The standard Otsu's and median extension methods diverged and gave erroneous segmentations. For the mixture-based methods, the MoGG gave a better segmentation than the MoG.

4.2. Simulated datasets

In this experiment, we used 5 benchmarks of randomly generated data. Each benchmark contains 100 datasets, and each dataset contains 10,000 data samples which correspond to gray levels. Each dataset is constituted of two classes ($\mathcal{C}_1$ for $r=1$ and $\mathcal{C}_2$ for $r=2$), where each class is modeled using a mixture of GGDs. The parameter setting of our generated benchmarks is presented in the Table 2. Our tests on the simulated datasets are conducted in two phases. In the first phase, the standard Otsu's and the median extension methods are compared with mixture-based methods (MoG and MoGG methods). In the second phase, we compared between the MoG and MoGG methods.

4.2.1. Otsu's based methods versus mixture methods

Conducted tests on our simulated datasets show that mixture-based models are very efficient against the standard Otsu's and the median-based extension methods. The key point of the mixture methods (MoG and MoGG) is that they consider the two classes (class $\mathcal{C}_{left}$ and class $\mathcal{C}_{right}$) follow a mixture of K_{left} and K_{right} components, respectively. In the case of the MoG model, distribution components are Gaussians and for the MoGG model, components are considered as generalized Gaussians. The standard Otsu's and Median-based extension methods consider that each class $\mathcal{C}_k$ follows a unimodal distribution, and therefore, fail to get the optimal threshold when the classes are multi-modal. To illustrate these facts, Fig. 7 shows two examples of generated datasets and thresholds obtained using the standard Otsu's, median-based extension, MoG and MoGG methods, respectively. Finally, Table 3 summarizes results obtained by the application of the four methods. It contains for each benchmark the average misclassification error (ME) of each method and, between brackets, the percentage of data sets where MoGG gave the least error. For the majority of the data sets, results show that MoGG method has a lesser ME than the other methods. This clearly demonstrates the performance of our approach.

4.2.2. MoG versus MoGG thresholding

This experiment aims to compare MoGG against MoG models for thresholding. Similarly to the experiment performed in [18] (see Chapter 1, page 12), three types of mixtures densities were generated, namely *bimodal*, *trimodal* and *multimodal* distributions, using GGD components instead of Gaussian distributions. Fig. 8 shows 4 examples of these histograms and the applied thresholding using MoG and MoGG models. These examples were chosen particularly because they contain modes with non-Gaussian shapes. The number of components K_{left} and K_{right} used for thresholding are those used in the sampling step. We can clearly note the improvement in terms of log-likelihood fitting by using MoGG method against the MoG method. In terms of thresholding, using MoG gave lesser performance than using MoGG in all the chosen examples. For instance, in the second example (top right), the right class (modeled by 2 Gaussians) was fitted completely to the rightmost mode which is very sharp, whereas

Table 6
Parameters setting for multimodal simulated dataset benchmarks.

Bench.	Class	K_r	μ	σ	β	π
Bench.1	$r=1$	2	$\mu_{1,1}=50$	$\sigma_{1,1}=20$	$\beta_{1,1}=2.00$	$\pi_{1,1}=0.40$
			$\mu_{1,2}=65$	$\sigma_{1,2}=20$	$\beta_{1,2}=2.00$	$\pi_{1,2}=0.20$
	$r=2$	2	$\mu_{2,1}=120$	$\sigma_{2,1}=20$	$\beta_{2,1}=3.00$	$\pi_{2,1}=0.15$
			$\mu_{2,2}=135$	$\sigma_{2,2}=20$	$\beta_{2,2}=1.50$	$\pi_{2,2}=0.15$
	$r=3$	1	$\mu_{3,1}=180$	$\sigma_{3,1}=15$	$\beta_{3,1}=1.00$	$\pi_{3,1}=0.10$
Bench.2	$r=1$	1	$\mu_{1,1}=4$	$\sigma_{1,1}=3$	$\beta_{1,1}=2.10$	$\pi_{1,1}=0.15$
	$r=2$	1	$\mu_{2,1}=22$	$\sigma_{2,1}=9$	$\beta_{2,1}=2.70$	$\pi_{2,1}=0.15$
	$r=3$	3	$\mu_{3,1}=77$	$\sigma_{3,1}=19$	$\beta_{3,1}=2.40$	$\pi_{3,1}=0.25$
			$\mu_{3,2}=107$	$\sigma_{3,2}=20$	$\beta_{3,2}=2.10$	$\pi_{3,2}=0.35$
			$\mu_{3,3}=133$	$\sigma_{3,3}=33$	$\beta_{3,3}=1.50$	$\pi_{3,3}=0.15$
Bench.3	$r=1$	2	$\mu_{1,1}=50$	$\sigma_{1,1}=10$	$\beta_{1,1}=2.00$	$\pi_{1,1}=0.30$
			$\mu_{1,2}=75$	$\sigma_{1,2}=15$	$\beta_{1,2}=2.00$	$\pi_{1,2}=0.15$
	$r=2$	2	$\mu_{2,1}=130$	$\sigma_{2,1}=25$	$\beta_{2,1}=3.10$	$\pi_{2,1}=0.15$
			$\mu_{2,2}=180$	$\sigma_{2,2}=15$	$\beta_{2,2}=1.30$	$\pi_{2,2}=0.30$
	$r=3$	1	$\mu_{3,1}=220$	$\sigma_{3,1}=5$	$\beta_{3,1}=2.30$	$\pi_{3,1}=0.10$
Bench.4	$r=1$	2	$\mu_{1,1}=85$	$\sigma_{1,1}=10$	$\beta_{1,1}=2.00$	$\pi_{1,1}=0.35$
			$\mu_{1,2}=100$	$\sigma_{1,2}=10$	$\beta_{1,2}=2.00$	$\pi_{1,2}=0.35$
	$r=2$	1	$\mu_{2,1}=150$	$\sigma_{2,1}=10$	$\beta_{2,1}=1.00$	$\pi_{2,1}=0.20$
	$r=3$	1	$\mu_{3,1}=180$	$\sigma_{3,1}=2$	$\beta_{3,1}=1.00$	$\pi_{3,1}=0.10$
Bench.5	$r=1$	2	$\mu_{1,1}=60$	$\sigma_{1,1}=15$	$\beta_{1,1}=4.00$	$\pi_{1,1}=0.20$
			$\mu_{1,2}=80$	$\sigma_{1,2}=15$	$\beta_{1,2}=4.00$	$\pi_{1,2}=0.20$
	$r=2$	1	$\mu_{2,1}=140$	$\sigma_{2,1}=15$	$\beta_{2,1}=1.00$	$\pi_{2,1}=0.30$
	$r=3$	1	$\mu_{3,1}=190$	$\sigma_{3,1}=8$	$\beta_{3,1}=1.00$	$\pi_{3,1}=0.30$

the MoGG used only one component for this mode, yielding finally to a nearly optimal threshold.

To emphasize our comparison, we generated 3 benchmarks of 100 datasets. Each dataset contains 10,000 data samples and is constituted of 2 classes generated using two MoGGs. Parameter setting of our generated datasets is presented in Table 4. Table 5 shows classification errors obtained using the MoG and MoGG models, respectively. We can note that in the 3 benchmarks, using MoGG gives better thresholds than using MoG. Figs. 9–11 show different examples where MoGG modeling was more efficient than using MoG's for thresholding. The third and fourth rows of each figure depict the histogram of each class ($\mathcal{C}_1$ and $\mathcal{C}_2$), the thresholds obtained by the application of MoG and MoGG methods and the plots of estimated mixture distribution fit of each class, respectively.

In Fig. 9, a two-class histogram is generated using the parameter setting of the benchmark 1 of Table 4 ($K_{left}=2$, $K_{right}=2$). We can observe that the MoGG method gave a better threshold $\simeq 101$ ($ME=0.0616$) against the MoG method which gives a

Table 7
ME for multi-level thresholding of the multimodal simulated datasets benchmarks (between brackets is the percentage of data sets where MoGG gives the least error).

Bench.	Error Otsu	Error Median	Error MoG	Error MoGG
1	0.050 (100%)	0.059 (100%)	0.034 (82%)	0.032
2	0.296 (100%)	0.269 (100%)	0.015 (60%)	0.015
3	0.159 (100%)	0.099 (100%)	0.026 (94%)	0.023
4	0.252 (100%)	0.263 (100%)	0.099 (94%)	0.023
5	0.015 (100%)	0.014 (98%)	0.151 (100%)	0.012

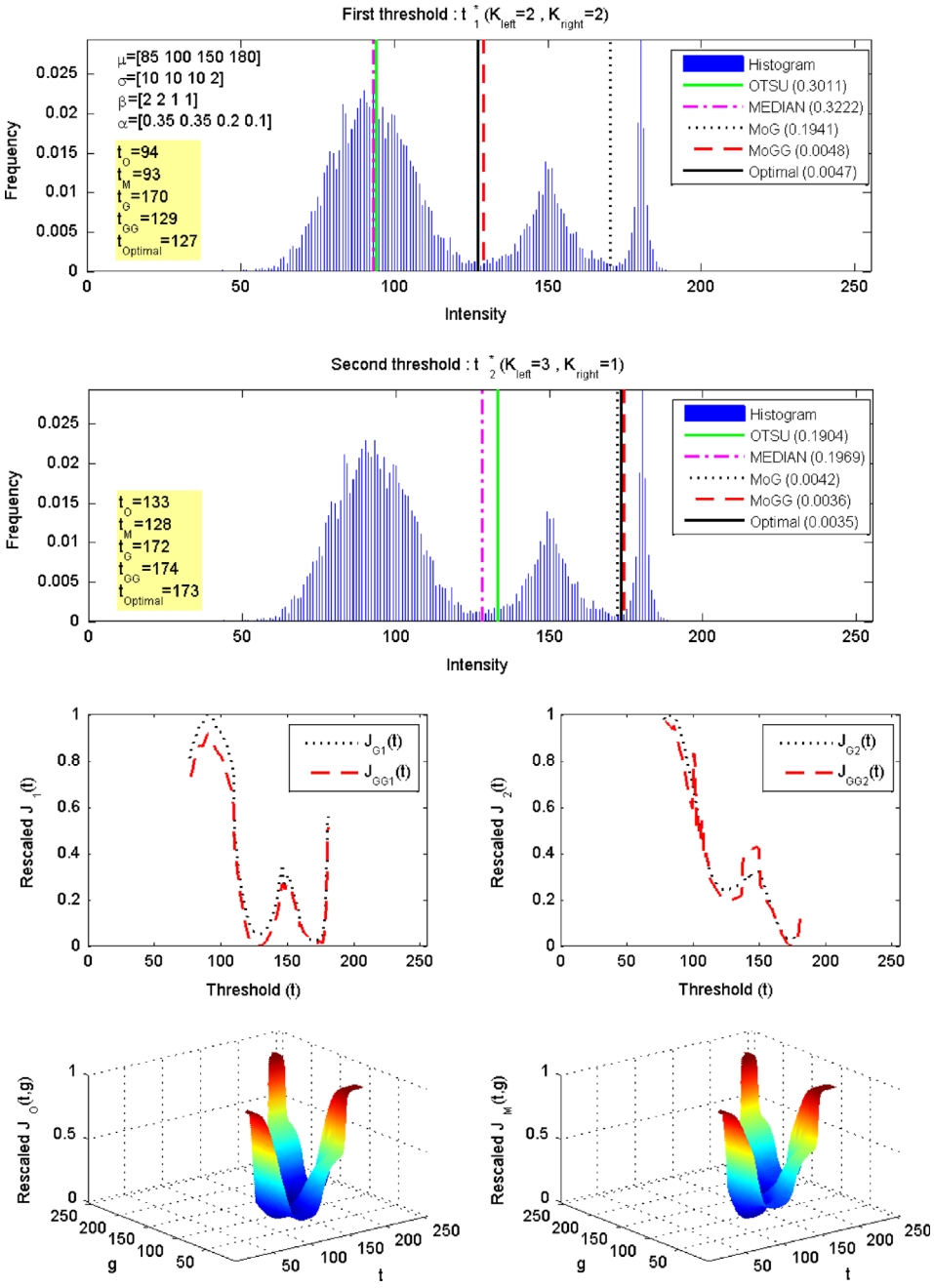

Fig. 12. Multi-thresholding results for a trimodal dataset histogram. From top to bottom and left to right, we show the first and second optimal thresholds ($\mathbf{t}_1^*$ and $\mathbf{t}_2^*$) given by all methods (first and second row), plots of MoG and MoGG criterion functions $J_1(\mathbf{t})$ and $J_2(\mathbf{t})$ respectively (third row) and plots of standard Otsu's method function $J_O(\mathbf{t},\mathbf{g})$ and median extension method function $J_M(\mathbf{t},\mathbf{g})$ (fourth row).

threshold $\simeq 128$ ($ME=0.2457$). This result is due to the non-Gaussian type of component distributions (see for instance the modes of the class in the right side). The right MoG model assigned its two components to the right mode. Consequently, it was not able to find the correct threshold between the classes $\mathcal{C}_1$ and $\mathcal{C}_2$ (see the graph on the bottom of the figure). The MoGG gave a nearly optimal threshold where the right MoGG model assigned its two components to the two rightmost modes of the histogram.

Another important fact that we have observed in the results, especially the benchmark 2, is the divergence of the MoG method for some examples. In Fig. 10, we show an example in benchmark 2 illustrating this fact. We can note that for the $J_{GG}(t)$ function, the global minimum is close to the optimal threshold $\simeq 100$ ($ME=0.0797$), whereas for the $J_G(t)$ function, the global minimum is far from the optimal threshold $\simeq 43$ ($ME=0.3458$). This problem of MoG-based thresholding is due to its inefficiency to fit accurately histograms which are constituted of non-Gaussian components (see the graph on the bottom of the figure). The MoGG model adequately fitted the non-Gaussian histogram modes in the left and right classes, and, consequently, gave a better threshold.

Finally, in Fig. 11, the left and right classes contain three and two modes, respectively. After computing the optimal thresholds using the two mixture models, the MoG method failed to place the optimal threshold in the true position $t_1 \simeq 177$ ($ME=0.2421$). In fact, two Gaussian components were necessary to fit one non-Gaussian histogram mode in the right class. The MoGG method successfully fitted each mode adequately with one mixture GGD component, giving, therefore, a better threshold $t_1 \simeq 123$ ($ME=0.012$). These examples demonstrate the performance of using MoGG models instead of MoG models for thresholding.

4.3. Multi-thresholding for simulated datasets ($K>2$)

Multi-thresholding has a variety of application fields (remote sensing, medical image segmentation, etc.). Similarly to the simple case of binarization ($K=2$), five benchmarks of randomly generated data were considered. Each benchmark contains 50 datasets with parameter settings presented in Table 6. For the case where the histogram is multimodal ($K>3$) and with different types of component shapes, results show that mixture-based methods (MoG and MoGG) are much more efficient compared to standard Otsu's and median extension methods.

Fig. 12 shows an example of tri-class histogram thresholding (i.e., $\mathbf{t}=(t_1^*,t_2^*)$). The minimized functions for the MoG and MoGG models are denoted by $J_G(\mathbf{t})=(J_{G1}(\mathbf{t}),J_{G2}(\mathbf{t}))$ and $J_{GG}(\mathbf{t})=(J_{GG1}(\mathbf{t}),J_{GG2}(\mathbf{t}))$, respectively. We can observe that the standard Otsu's and median based methods failed to find good thresholds because of the difference between shape and size parameters of the three classes ($\mathcal{C}_1,\mathcal{C}_2,\mathcal{C}_3$): (i) for the scale parameter σ, $\mathcal{C}_1$ and $\mathcal{C}_2$ have $\sigma_1=10$ and $\sigma_2=10$, respectively, while the only component that constitute $\mathcal{C}_3$ class have $\sigma_3=2$, and (ii) the proportion parameter π: $\mathcal{C}_1$ have each one $\pi_{11}=\pi_{12}=0.35$ for the class $\mathcal{C}_2$ have $\pi_2=0.20$, whereas $\mathcal{C}_3$ has a value of $\pi_3=0.10$. The MoG method failed to find the first threshold and succeeded in finding the second one, whereas the MoGG method succeeded in finding both thresholds.

Table 7 summarizes results obtained by all studied methods for all generated data sets. It shows for each benchmark average ME values obtained by each method and between brackets the percentage of data sets where MoGG method gave the best threshold. We note that for the case of three-levels thresholding, the misclassification error (ME) is computed as the average of the two thresholds t_1^* and t_2^* computed ME (i.e., $ME=(ME_1+ME_2)/2$), where ME_1 and ME_2 are values of ME corresponding to thresholds t_1^* and t_2^*, respectively.

4.4. Multi-thresholding for real images ($K>2$)

Here, two sample images were used, the 'Lake' (512×512) classic benchmark test image that was also employed in [30], and another image from the web. The second image is characterized by

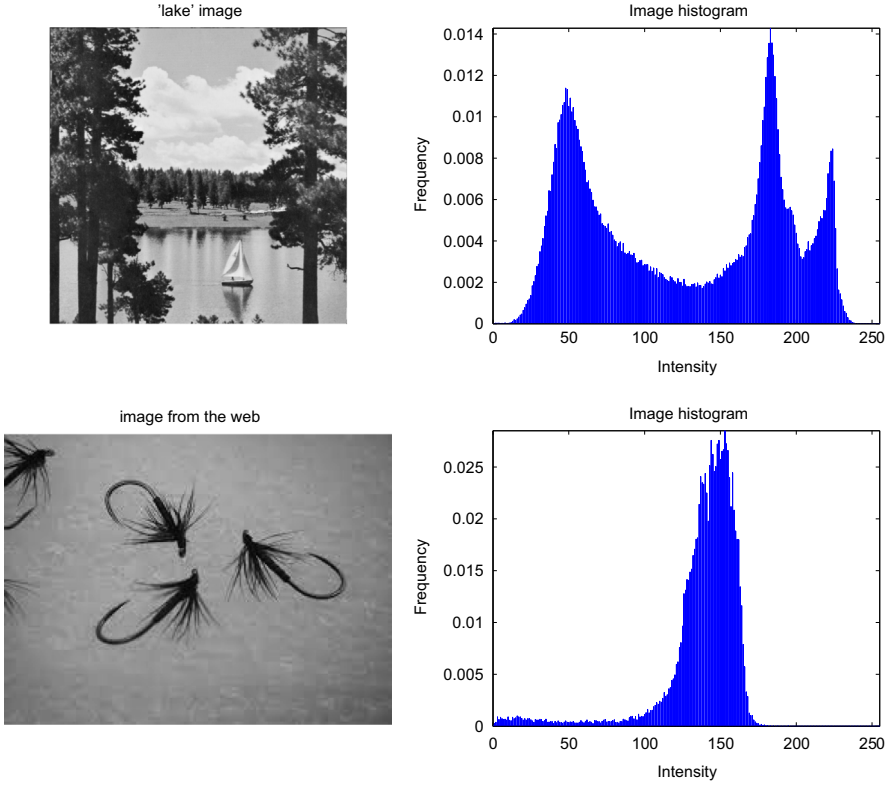

Fig. 13. Real images for multi-thresholding tests. From top to down, left to right, we show the 'Lake' original image, its histogram (first row) and an NDT–image 20 and its histogram (second row).

a non-uniform background. For both images, the parameters K_1, K_2 and K_3 represent the number of components in each class $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$, respectively (see Fig. 13).

Fig. 14 illustrates the multi-thresholding results for the 'Lake' image. In the shown histogram, there are three clearly perceptible modes and one or more classes with skewed distributions. We can observe that only mixture based methods successfully identified the three classes represented by the three modes of intensity (see Figs. 14 and 15). Although the first optimal threshold t_1 can be determined successfully by all methods, both Otsu's based methods failed to place the second optimal threshold t_2 in the right position of the histogram. This is mainly due to:

(i) the difference between the parameters π_1, π_2 and π_3 that represents the size of classes $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$, respectively. It can be observed that there is a great difference between π_1 and π_2 against π_3, (ii) the skewness of components distributions. The same comments can be made about Fig. 16 for the other image, where the background of the image was over-segmented using the standard Otsu's and median extension methods, as well as the MoG-based method. The MoGG gave a better segmentation compared to the other methods since it was able to model successfully the non-uniform background in one single GGD (see Figs. 16 and 17).

5. Computational time complexity

The computational complexity of the standard Otsu's and its median extension approaches grows exponentially with the number of thresholds and gray levels. The two methods can compute

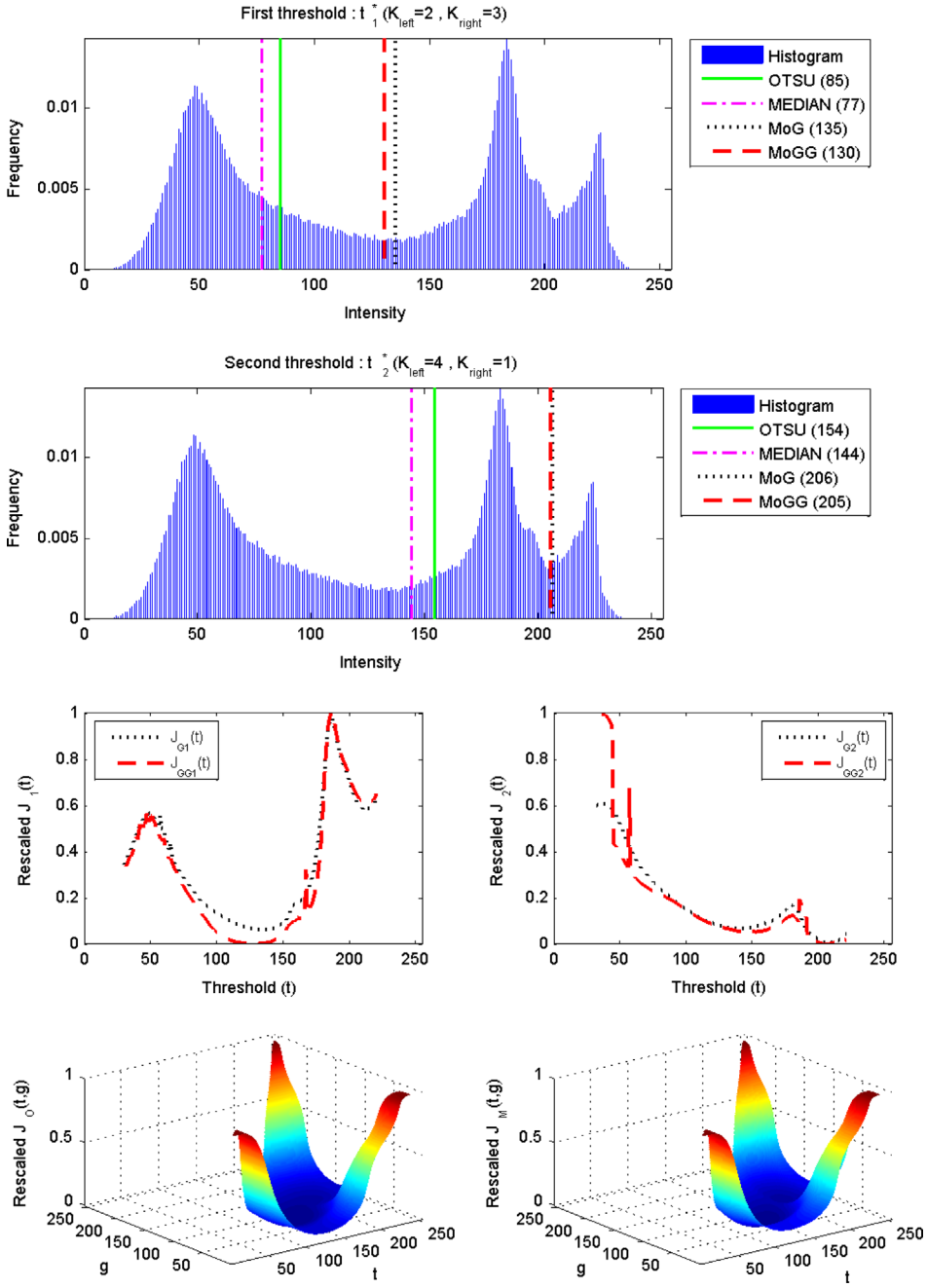

Fig. 14. Multi-thresholding results for 'Lake' image. From top to down, left to right, we show the first and second optimal thresholds ($\mathbf{t}_1^*$ and $\mathbf{t}_2^*$) given by all methods (first and second row), plots of MoG and MoGG criterion functions $J_1(\mathbf{t})$ and $J_2(\mathbf{t})$ respectively (third row) and plots of standard Otsu's method function $J_O(\mathbf{t},\mathbf{g})$ and median extension method function $J_M(\mathbf{t},\mathbf{g})$ (fourth row).

Fig. 15. Segmentation results for the 'lake' image obtained by application of standard Otsu's, median based extension, MoG and MoGG methods, respectively. (a) Standard Otsu's method, (b) Median based extension method, (c) MoG model and (d) MoGG model.

the optimal threshold in $\mathcal{O}(N^L)$ operations, where N is the number of gray levels, $L=K-1$ is the number of thresholds and K is the number of mixture components. This complexity is due to the minimization of criterion functions $J_O(\mathbf{t})$ and $J_M(\mathbf{t})$ which have L-dimensions (i.e., a function of L variables). This property considerably increases the computation time, especially for large values of K (eg. $K>5$). Several works in the past (e.g., see [17,33]) proposed faster versions for multi-level thresholding using the Otsu's method.

In our approach, the thresholding is based on fitting a mixture of generalized Gaussian distributions to multi-modal classes, where the EM algorithm is used to estimate the mixture parameters. Consequently, the MoGG method adds some computation time for determining the thresholds, compared to the standard Otsu's and its median extension variant. However, given the improvement of accuracy in the thresholds determination, this additional time is not a limitation. To speed up threshold calculations, the function $J_{GG}(\mathbf{t})$ is rearranged to be uni-dimensional (i.e,. a function of one variable) whatever the number of classes K; thus, the proposed MoGG method has a linear complexity. In fact, the proposed algorithm can find optimal thresholds in time $\mathcal{O}(NL)$ operations of EM algorithm estimation (see Algorithm 1). Consequently, for high values of K (e.g., $K>5$), the proposed mixture methods consume less time than standard Otsu's and median extension methods.

Figs. 14, 16 and 12 show examples of multi-thresholding ($K=3$), where we can observe that $J_G(\mathbf{t})$ and $J_{GG}(\mathbf{t})$ functions are uni-dimensional while $J_O(\mathbf{t},\mathbf{g})$ and $J_M(\mathbf{t},\mathbf{g})$ are two-dimensional (i.e., they are defined by two variables $\mathbf{t}$ and $\mathbf{g}$).

6. Remarks and discussion

In this section, we present some remarks and caveats about the approaches studied in this paper, and thresholding-based segmentation in general. These revolve around the multi-modality of criterion function used for threshold determination. We observed in some cases that the best threshold is not given by the global minimum but by a local minimum. This limitation was already observed for the standard Otsu's method in [13]. It has been demonstrated that the objective function may not only be multi-modal, but more importantly, if it is multi-modal, its global maximum is not guaranteed to give a correct threshold. For the median-based extension [30], the authors used a two-component Laplace mixture to simulate data, where the two components (or classes) have greatly disparate sizes (with a ratio of 99–1). They observed that sometimes for the standard Otsu's method and its median based extension, local minima of $J(\mathbf{t})$ provide better thresholds than do global minima.

To illustrate the behavior of MoG and MoGG criterion function against Otsu's and median extension ones, we conducted the same experiments as in [13] and [30]. Two datasets were randomly generated, each of which is a two-component Laplace and Gaussian mixture, respectively, where the two classes have the following parameter settings. For the left class, we used $\mu_1=80$, $\sigma_1=15$, $\pi_1=0.99$, while for the right class, we used $\mu_2=190$, $\sigma_2=20$, $\pi_2=0.01$. These parameters guarantee an inter-class ratio of 99–1. Figs. 18 and 19 show thresholding results and multi-modal criterion functions ($J_O(\mathbf{t})$, $J_M(\mathbf{t})$, $J_G(\mathbf{t})$ and $J_{GG}(\mathbf{t})$) obtained by all methods for the two cases of Gaussian and Laplace datasets.

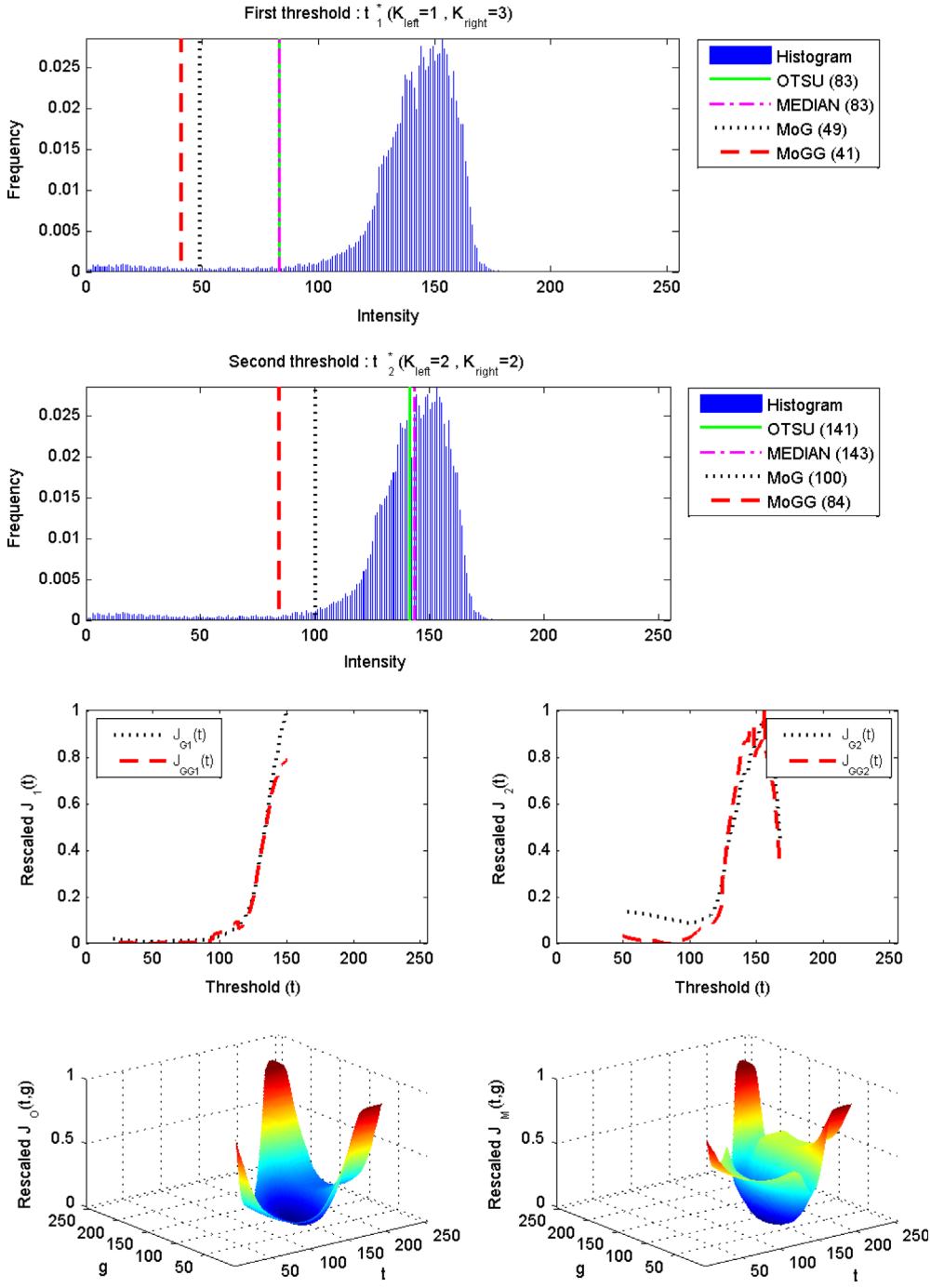

Fig. 16. Multi-thresholding results for the second image. Caption is as for Fig. 14.

For standard Otsu's and median extension methods, the local minima at 150 in Fig. 18 gives better thresholds. Functions $J_O(\mathbf{t})$ and $J_M(\mathbf{t})$ are multi-modal and the local minimum indicates a better threshold than does the global minimum. Consequently, the two methods fail both to find a good threshold. However, for the mixture-based methods (MoG and MoGG), despite the multi-modality of criterion functions $J_G(\mathbf{t})$ and $J_{GG}(\mathbf{t})$, the two methods give good thresholds and successfully separate the two classes $\mathcal{C}_1$ and $\mathcal{C}_2$. In Fig. 19, the fact that MoGG method gives better threshold (at about 162) than that given by MoG method (at about 130) is due to the shape parameter β which allows MoGG method to fit the Laplace distribution better than the MoG method.

We note finally that to resolve the problem of multi-modality of the criterion function, a simple valley check can be used to enable the thresholding method to be extended to reliably estimate thresholds over a larger range of inter-class size ratios [13]. For instance, a threshold t^* that does not satisfy $h(t^*) < h(\overline{x}_1(t^*))$ and $h(t^*) < h(\overline{x}_2(t^*))$ will be ignored. For MoG and MoGG multi-class thresholding methods, we propose the following "valley check" : let $\overline{x}$ the mixture mean vector estimated by EM algorithm, for $r = \{1, \ldots, K-1\}$ a threshold t_r^* must satisfy

$$\overline{x}\left(\sum_{i=1}^{r} K_i\right) < t_r^* < \overline{x}\left(\sum_{i=1}^{r+1} K_i\right), \tag{27}$$

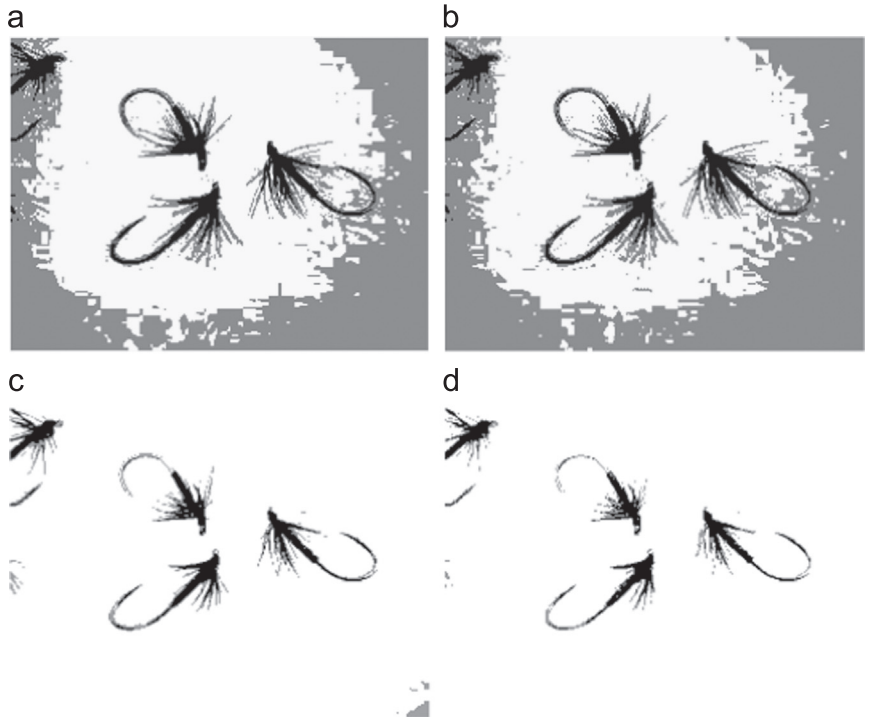

Fig. 17. Segmentation results of the second image by application of standard Otsu's, median based extension , MoG and MoGG methods, respectively. (a) Standard Otsu's method, (b) Median based extension method, (c) MoG model and (d) MoGG model.

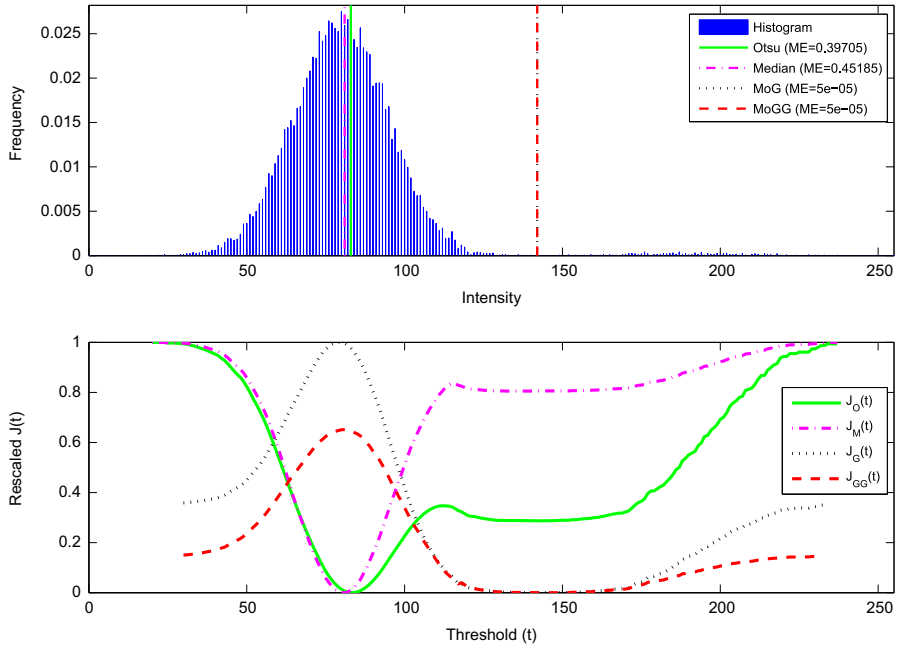

Fig. 18. Thresholds and criterion function for the Otsu's, the median-based extension, MoG and MoGG methods applied for segmenting data simulated from mixtures of two Gaussian distributions.

where K_i is the number of mixture components in class C_i (see Algorithm 1).

7. Conclusion

A new thresholding approach, based on the Mixture of Generalized Gaussian model (MoGG method), is presented in this paper. The approach generalizes previous methods to multi-thresholding and multi-modal classes. It has been successfully tested on segmentation of real images (NDT-images [27]) and randomly generated data sets. Experiments have shown the performance of the proposed approach and showed that it can achieve more optimal thresholds than the standard Otsu's method [22], the median based extension method [30], as well as thresholding based on Gaussian mixture models.

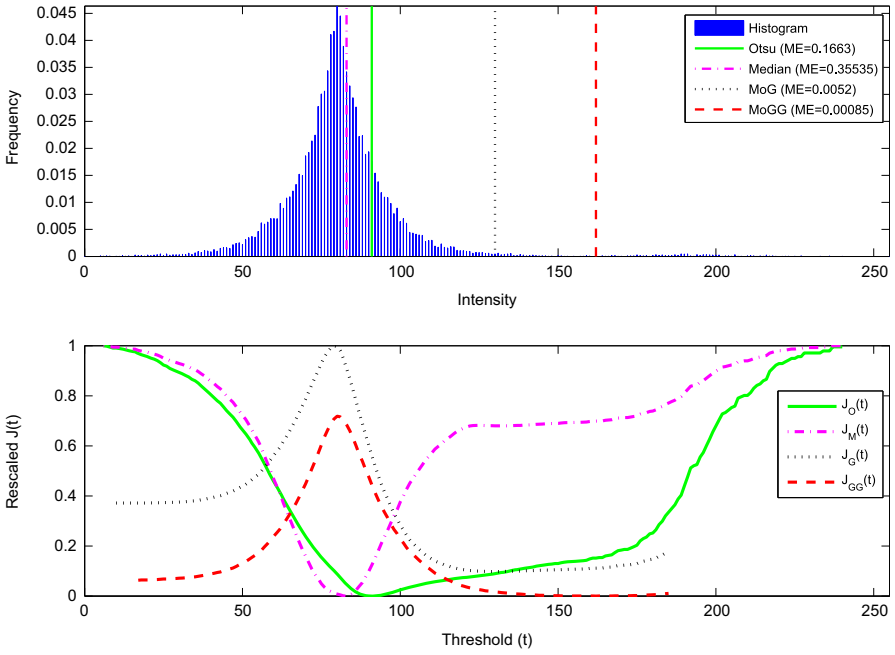

Fig. 19. Thresholds and criterion function for the Otsu's, the median-based extension, MoG and MoGG methods applied for segmenting data simulated from mixtures of two Laplace distributions.

Conflict of Interest statement

None declared.

References

[1] M.S. Allili, D. Ziou, N. Bouguila, Finite General Gaussian mixture modelling and application to image and video foreground segmentation, Journal of Electronic Imaging 17 (1) (2008) 001–013.

[2] Y. Bazi, L. Bruzzone, F. Melgani, Image thresholding based on the EM algorithm and the generalized Gaussian distribution, Pattern Recognition 40 (2) (2008) 619–634.

[3] B. Bhanu, J. Peng, Adaptive Integrated Image Segmentation and Object Recognition, IEEE Transactions on Systems, Man, and Cybernetics—Part C 30 (4) (2000) 427–441.

[4] G-E.P. Box, G.C. Tiao, Bayesian Inference in Statistical Analysis, Wiley Classics Library, 1992.

[5] A.D. Brink, N.E. Pendock, Minimum cross entropy threshold selection, Pattern Recognition 29 (1) (1996) 179–188.

[6] C.-I. Chang, Y. Du, J. Wang, S.-M. Guo, P.D. Thouin, Survey and comparative analysis of entropy and relative entropy thresholding techniques, IEE Proceedings – Vision Image and Signal Processing 153 (6) (2006) 837–850.

[7] S.-K.S. Fan, Y. Lin, C.-C. Wu, Image thresholding using a novel estimation method in generalised Gaussian distribution mixture modelling, Neurocomputing 72 (1–3) (2008) 500–512.

[8] S.-K.S. Fan, Y. Lin, A fast estimation method for the generalized Gaussian mixture distribution on complex images, Computer Vision and Image Understanding 113 (7) (2009) 839–853.

[9] M. Figueiredo, A.K. Jain, Unsupervised learning of finite mixture models, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (3) (2002) 381–396.

[10] C.A. Glasbey, An analysis of histogram-based thresholding algorithms, CVGIP: Graphical Models and Image Processing 55 (6) (1993) 532–537.

[11] P.J. Huber, E.M. Ronchetti, Robust Statistics, 2nd Edition, Wiley Series in Probability and Statistics, 2009.

[12] F. Jiulun, X. Winxin, Minimum error thresholding: a note, Pattern Recognition Letters 18 (8) (1997) 705–709.

[13] J. Kittler, J. Illingworth, On threshold selection using clustering criteria, IEEE Transactions on Systems, Man, and Cybernetics 15 (5) (1985) 652–655.

[14] J. Kittler, J. Illingworth, Minimum error thresholding, Pattern Recognition 19 (1) (1986) 41–47.

[15] S. Kullback, Information Theory and Statistics, Dover, 1968.

[16] T. Kurita, N. Otsu, N. Abdelmalek, Maximum likelihood thresholding based on population mixture models, Pattern Recognition 25 (10) (1992) 1231–1240.

[17] P.-S. Liao, T.-S. Chen, P.-C. Chung, Multi-level thresholding for image segmentation through a fast statistical recursive algorithm, Journal of Information Science and Engineering 17 (5) (2001) 713–727.

[18] G. McLachlan, D. Peel, Finite Mixture Models, Wiley Series in Probability and Statistics, 2000.

[19] F. Morii, A note on minimum error thresholding, Pattern Recognition Letters 12 (6) (1991) 349–351.

[20] G. Moser, B. Serpico, Generalized minimum-error thresholding for unsupervised change detection from SAR amplitude images, IEEE Transactions on Geoscience and Remote Sensing 44 (10) (2006) 2972–2982.

[21] N. Nacereddine, S. Tabbone, D. Ziou, L. Hamami, Asymmetric generalized Gaussian mixture models and EM algorithm for image segmentation, in: International Conference on Pattern Recognition, 2010, pp. 4557–4560.

[22] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics 9 (1) (1979) 62–66.

[23] N.R. Pal, D. Bhandari, Image thresholding: some new techniques, Signal Processing 33 (2) (1993) 139–158.

[24] P.K. Sahoo, S. Soltani, A.K.C. Wong, Y.C. Chen, A survey of thresholding techniques, Computer Vision, Graphics, and Image Processing 41 (2) (1988) 233–260.

[25] P.K. Sahoo, G. Arora, Image thresholding using two-dimensional Tsallis–Havrda–Charvát entropy, Pattern Recognition Letters 27 (6) (2006) 520–528.

[26] A.E. Savakis, Adaptive document image thresholding using foreground and background clustering, IEEE International Conference on Image Processing (3) (1998) 785–789.

[27] M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, Journal of Electronic Imaging 13 (1) (2004) 146–165.

[28] S. Wang, F.-L. Chung, F. Xionga, A novel image thresholding method based on Parzen window estimate, Pattern Recognition 41 (1) (2008) 117–129.

[29] J.-H. Xue, Y.-J. Zhang, X.G. Lin, Rayleigh-distribution based minimum error thresholding for SAR images, Journal of Electronics (China) 16 (4) (1999) 336–342.

[30] J.-H. Xue, D.M. Titterington, Median-based image thresholding, Image and Vision Computing 29 (9) (2011) 631–637.

[31] J.-H. Xue, D.M. Titterington, Threshold selection from image histograms with skewed components based on maximum-likelihood estimation of skewnormal and log-concave distributions, Id 7347, oai:eprints.pascal-network.org:7347, 2011.

[32] J.-H. Xue, Y.-J. Zhang, Ridler and Calvard's, Kittler and Illingworth's and Otsu's methods for image thresholding, Pattern Recognition Letters 33 (6) (2012) 793–797.

[33] P.-Y. Yin, L.-H. Chen, A fast iterative scheme for multilevel thresholding methods, Signal Processing 60 (3) (1997) 305–313.

[34] S.X. Yu, R. Gross, J. Shi, Concurrent object recognition and segmentation by graph partitioning, Neural Information Processing Systems (2002) 1383–1390.

[35] M.S. Allili, D. Ziou, N. Bouguila, S. Boutemedjet, Image and Video Segmentation by Combining Unsupervised Generalized Gaussian Mixture Modeling and Feature Selection. IEEE Transaction on Circuits and Systems for Video Technology 20 (10) (2010) 1373–1377.

[36] M.S. Allili, D. Ziou, Globally adaptive region information for automatic color-texture image segmentation. Pattern Recognition Letters 28 (15) (2007) 1946–1956.

Aïssa Boulmerka received the B Eng and MSc degrees in computer science from École nationale Supérieure en Informatique (Algeria) in 2004 and 2009, respectively. Since 2010, he has been pursuing PhD studies at École nationale Supérieure en Informatique (Algeria) and Université du Québec in Outaouais (Canada) under the supervision of Professor Samy Ait-Aoudia and Professor Mohand Saïd Allili. His primary research interests include statistical models and application to image segmentation, computer vision and pattern recognition.

Mohand Saïd Allili received the M.Sc. and Ph.D. degrees in computer science from the University of Sherbrooke, Sherbrooke, QC, Canada, in 2004 and 2008, respectively. Since June 2008, he has been an Assistant Professor of computer science with the Department of Computer Science and Engineering, Université du Québec en Outaouais, Canada. His main research interests include computer vision and graphics, image processing, pattern recognition, and machine learning. Dr. Allili was a recipient of the Best Ph.D. Thesis Award in engineering and natural sciences from the University of Sherbrooke for 2008 and the Best Student Paper and Best Vision Paper awards for two of his papers at the Canadian Conference on Computer and Robot Vision 2007 and 2010, respectively.

Samy Ait-Aoudia received a DEA "Diplôme d'Etudes Approfondies" in image processing from Saint-Etienne University, France, in 1990. He holds a Ph.D. degree in computer science from the Ecole des Mines, Saint-Etienne, France, in 1994. He is currently a Professor of computer science at the National High School in Computer Science at Algiers/Algeria, where he is involved in teaching BSc and MSc levels in computer science and software engineering. His areas of research include image processing, CAD/CAM and constraints management in solid modeling.