

# Image thresholding based on the EM algorithm and the generalized Gaussian distribution

Yakoub Bazi, Lorenzo Bruzzone\*, Farid Melgani

*Department of Information and Communication Technologies, University of Trento, Via Sommarive, 14, I-38050, Trento, Italy*

Received 16 August 2005; accepted 9 May 2006

---

## Abstract

In this paper, a novel parametric and global image histogram thresholding method is presented. It is based on the estimation of the statistical parameters of “object” and “background” classes by the expectation–maximization (EM) algorithm, under the assumption that these two classes follow a generalized Gaussian (GG) distribution. The adoption of such a statistical model as an alternative to the more common Gaussian model is motivated by its attractive capability to approximate a broad variety of statistical behaviors with a small number of parameters. Since the quality of the solution provided by the iterative EM algorithm is strongly affected by initial conditions (which, if inappropriately set, may lead to unreliable estimation), a robust initialization strategy based on genetic algorithms (GAs) is proposed. Experimental results obtained on simulated and real images confirm the effectiveness of the proposed method.

© 2006 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

*Keywords:* Image thresholding; Expectation–Maximization algorithm; Generalized Gaussian distribution; Genetic algorithms

---

## 1. Introduction

Thresholding is a fundamental task in image processing and pattern recognition as it represents a basic step for image understanding in numerous applications. It consists of converting gray-level images into binary images by selecting an appropriate decision threshold. The gray-level values below or equal to the selected threshold are usually classified as background, while the values above this threshold are classified as object (or vice versa). Image thresholding is widely used in many application domains, such as biomedical image analysis [1], handwritten character identification [2], automatic target recognition [3], change-detection applications [4–6], etc.

In general, automatic thresholding techniques are classified into two main groups: global methods and local methods. Global methods adopt a fixed threshold for the entire image, while local methods exploit threshold values that change dynamically in the image. Local methods are

generally used when background and object classes have statistical properties that are not stationary in the different portions of the analyzed image (for example, due to nonhomogeneous light conditions) (see, for example, Refs. [7,8]). In this paper, we address the former group, i.e. global thresholding methods.

Many techniques for global image thresholding have been proposed in the literature [9–22]. They are generally based on parametric or nonparametric approaches. In the parametric approach, object and background classes are discriminated (1) by assuming a predefined statistical model for approximating the class distributions and (2) by determining the optimal threshold value as a function of the statistical parameters of the two classes. By contrast, the nonparametric approach does not make any assumption about the class statistical distributions. Among the early global thresholding techniques available in the literature, one can find the work presented in Refs. [9,10]. In this iterative method, the threshold value is first initialized with the mean of the entire histogram. Then at each iteration the new threshold value is computed as the average of the two mean values of the two distributions on the left and right-hand side of the

---

\* Corresponding author. Tel.: +39 0461 882056; fax: +39 0461 882093.  
E-mail address: [lorenzo.bruzzone@ing.unitn.it](mailto:lorenzo.bruzzone@ing.unitn.it) (L. Bruzzone).

threshold estimated at the previous iteration. This iterative process continues until convergence is reached. In Ref. [11], the authors propose to select the optimal threshold value through a discriminant criterion that defines a separability measure between the object and background classes on the basis of second-order statistics. In Ref. [12], a method based on the moment preserving principle is presented, in which the threshold value is selected in such a way that the first three moments of the original image are preserved in the binary image. In Ref. [13], the threshold value is identified by maximizing the sum of entropies of object and background classes of the image. In Ref. [14], a method that optimizes a criterion function based on the Bayes classification rule for minimum error and on the Gaussian assumption for object and background classes is presented. In Ref. [15], the optimal threshold is determined by maximizing the posterior entropy subject to inequality constraints derived from measures of the uniformity and shape of the regions in the image. In Ref. [16], an extension of the methods presented in Refs. [13,15] using 2-D entropies is proposed. The 2-D entropies are obtained from a bi-dimensional histogram constructed using the gray-level values and the local average gray values. In Ref. [17], a maximum likelihood (ML) thresholding method based on a population of Gaussian mixtures is introduced. In this work, the authors show that the maximization of the likelihood of the conditional distributions under the Gaussian assumption with equal and different variances is equivalent to the methods developed in Refs. [11,14], respectively. In Ref. [18], a technique based on Kullback's minimum cross-entropy principle is described. The optimum threshold is obtained by minimizing the aforementioned cross entropy (formulated on a pixel-to-pixel basis) between the object and background classes. An algorithm that maximizes the sum of the entropies computed from the autocorrelation functions of the thresholded image histogram is presented in Ref. [19]. More recently, other global thresholding procedures based on fuzzy logic have been introduced [20–22]. Their underlying idea is to determine the best threshold value by minimizing a measure of fuzziness (ambiguity) of an image. Numerous other image thresholding methods can be found in the image processing and pattern recognition literature. Among others the interesting comparative studies reported in Refs. [6,23–26] deserve to be pointed out.

Usually, thresholding techniques based on parametric models (e.g., those developed in Refs. [9,10,14,17]) suffer from some serious drawbacks (which may be very critical in complex images) such as (i) they are based on the Gaussian assumption for modeling the class distributions in the image, which often does not hold in real images; (ii) they often lead to biased estimates of the statistical parameters of the object and background classes; and (iii) their effectiveness is strongly reduced when the prior probabilities of object and background classes are unbalanced or when the two classes overlap significantly [23–26].

In order to address these drawbacks, in this paper, we propose a novel parametric and global image histogram thresholding technique. It is based on the expectation–maximization (EM) algorithm to estimate the parameters of the object and background classes, which are assumed to follow a generalized Gaussian (GG) distribution. In the literature, the EM algorithm has already been used for global image histogram thresholding [27] and for the complex problem of change detection in multitemporal remote-sensing (RS) images [4] by assuming that the two above classes follow a Gaussian distribution. In order to cope with the limitations of the Gaussian model, an approach combining the EM algorithm with a semi-parametric model based on the Parzen window was proposed in Ref. [5]. In spite of the fact that this approach has proved effective in many thresholding problems for complex cases, it may exhibit problems of stability due: (i) to the relatively large number of parameters to be estimated in an unsupervised way and (ii) to the empirical procedure used to initialize the EM algorithm. In this paper, we propose to use the GG model as an alternative model for approximating the aforementioned classes. Compared to the semi-parametric model, the GG model is intrinsically more stable, since it is characterized by fewer parameters to be estimated (three parameters for each class). Compared to the Gaussian model, thanks to an additional statistical parameter (i.e., the shape parameter), it is more flexible and can approximate a large class of statistical distributions. After formulating the thresholding problem as a binary classification problem with a mixture of GG distributions, the iterative EM equations for the parameter estimation of the GG mixture are derived. Since the goodness of the solution found at convergence by the EM algorithm strongly depends on the adopted initialization procedure, we propose to identify the optimal set of initial parameter values by means of a robust initialization strategy based on genetic algorithms (GAs).

The paper is organized in six sections. The problem formulation is reported in Section 2. A detailed description of the proposed method is introduced in Section 3. In Section 4, the proposed initialization procedure based on GAs is presented. Descriptions of the data sets and the experiments are given in Section 5. Finally, conclusions and discussions are drawn in Section 6.

## 2. Problem formulation

Let us consider an image  $X$  of size  $M \times N$  pixels, where each pixel  $x$  of coordinates  $(m, n) \in [1, M] \times [1, N]$  can take  $L$  possible gray-level values defined in the range  $[0, L - 1]$ . Let  $h(x)$  be the normalized histogram of the image  $X$  (we assume that  $h(x)$  is the only information available about the image). The histogram  $h(x)$  can be seen as an estimate of the true probability density function  $p(x)$  of the image. We aim at thresholding the image  $X$  in order to assign each image pixel to one of the two opposite classes, namely object

and background. This task can be reduced to a problem of selection of the threshold value  $T$  that optimizes a predefined criterion. Once  $T$  is computed, the thresholded image  $Y = \{y(m, n), 1 \leq m \leq M, 1 \leq n \leq N\}$  can be generated by assigning the following values:

$$y(m, n) = \begin{cases} 0 & \text{if } x(m, n) \leq T, \\ 255 & \text{if } x(m, n) > T. \end{cases} \quad (1)$$

The above problem can be solved as a binary classification problem, where the probability density function  $p(x)$  of the image  $X$  is a mixture of two parametric density functions associated with the object and background classes, i.e.,

$$p(x) = \sum_{i=1}^2 P_i p_i(x|\theta_i), \quad (2)$$

where  $P_1$  and  $P_2$  are the prior probabilities, and  $p_1(x|\theta_1)$  and  $p_2(x|\theta_2)$  the class-conditional densities associated with the object and background classes, respectively.  $\theta_1$  and  $\theta_2$  are the vectors of parameters on which the two parametric class-conditional densities depend. The Generation of the binary classification map  $Y$  requires: (1) the adoption of a statistical model for the two class-conditional densities; (2) an estimation of the statistical parameters related to the two classes (i.e.,  $P_i$  and  $\theta_i$ , with  $i = 1, 2$ ); and (3) the application of a decision criterion.

In this paper, we propose to solve this problem through a formulation of the EM algorithm under the assumption that the class-conditional densities  $p_1(x|\theta_1)$  and  $p_2(x|\theta_2)$  follow a GG distribution. The thresholded image  $Y$  is then generated by exploiting the estimated parameters in the context of a Bayesian decision rule (e.g., minimum error, minimum cost, etc.).

### 3. EM algorithm under the GG assumption

#### 3.1. GG distribution

In order to cope with the limitations of the Gaussian model, we need to use another model that can describe in the best possible way the statistical behaviors of the object and background classes in a generic image. A possible solution is to adopt a more general parametric model that should satisfy two main properties: (i) flexibility (i.e., it should be capable of modeling a large variety of statistical behaviors) and (ii) stability (i.e., it should not require the estimation of a large number of parameters). Among the models available in the literature, the GG distribution is a particularly attractive candidate as it requires only one additional parameter to be estimated compared to the Gaussian model, and it can approximate a large class of statistical distributions (e.g., impulsive, Laplacian, Gaussian, and uniform distributions). The analytical equation of the GG distribution we adopt in this paper for modeling the two class-conditional densities

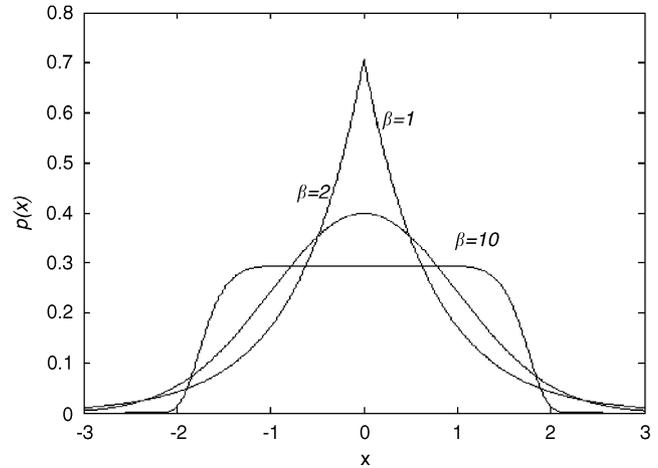


Fig. 1. GG distribution plotted for different shape parameters in the assumption of zero mean and scale parameter equal to one.

is given by [28]

$$p_i(x|\theta_i) = \frac{\beta_i}{2\alpha_i \Gamma(1/\beta_i)} e^{-[|x-\mu_i|/\alpha_i]^{\beta_i}}, \quad i = 1, 2, \quad (3)$$

where  $\mu_i$ ,  $\alpha_i$  and  $\beta_i$  are the mean, the scale, and the shape parameters of the  $i$ th class-conditional distribution, respectively, and  $\Gamma(\cdot)$  is the Gamma function, which is defined as  $\Gamma(v) = \int_0^\infty e^{-t} t^{v-1} dt$ . The scale parameter that expresses the width of the distribution is related to the normal standard deviation by the equation:

$$\alpha_i = \sigma_i \sqrt{\frac{\Gamma(1/\beta_i)}{\Gamma(3/\beta_i)}}. \quad (4)$$

The shape parameter  $\beta_i$  tunes the decay rate of the density function. It is worth noting that  $\beta_i = 2$  yields the Gaussian (normal) density and  $\beta_i = 1$  results in the Laplacian density function. As limit cases, for  $\beta_i \rightarrow 0$  the distribution becomes impulsive, whereas for  $\beta_i \rightarrow \infty$  it approaches a uniform distribution. Fig. 1 shows the plot of GG distributions for different values of the shape parameter  $\beta_i$  in the assumption of zero mean and scale parameter equal to one.

#### 3.2. EM estimation algorithm for the GG distribution

A common way of estimating the set of statistical parameters  $\theta = [P_1, P_2, \theta_1, \theta_2]$  related to the two classes object and background consists of maximizing the log likelihood  $L(X|\theta)$  of the two-component mixture defined in (2), i.e.,

$$\theta^* = \arg \max_{\theta} \{L(X|\theta)\}, \quad (5)$$

where

$$L(X, \theta) = \sum_{x=0}^{L-1} h(x) \ln[p(x|\theta)]. \quad (6)$$

It is well known that the solution to the above problem cannot be found analytically, as it is an intrinsically ill-posed problem (many-to-one mapping) that can result in different solutions. A possible choice to obtain ML estimates of the mixture parameters is to use the EM algorithm [29]. It is based on the interpretation of  $X$  as incomplete data, where the missing part is  $Y$  (i.e., the thresholded image). In other words, the missing part can be evaluated as a set of  $L$  labels  $Z = \{z(x) : x \in [0, L - 1]\}$  associated with the  $L$  possible sample realizations  $x \in [0, L - 1]$ . Each label  $z(x)$  of  $Z$  indicates which component (i.e., object or background) is at the origin of the realization  $x$ . It is a binary vector given by  $z(x) = \{z_1(x), z_2(x)\}$  with  $z_i(x) = 1$  if  $x$  belongs to the component  $i$  and  $z_i(x) = 0$  otherwise. The complete data log likelihood function becomes [30]:

$$L(X, Z, \theta) = \sum_{i=1}^2 \sum_{x=0}^{L-1} z_i(x) h(x) \ln[P_i p_i(x|\theta_i)], \quad (7)$$

where  $\theta_i = [\mu_i, \alpha_i, \beta_i]$ .

In the case of a GG distribution, it is possible to prove that the complete log likelihood is given by

$$\begin{aligned} L(X, Z, \theta) &= \sum_{i=1}^2 \sum_{x=0}^{L-1} z_i(x) h(x) \ln P_i + \sum_{i=1}^2 \sum_{x=0}^{L-1} z_i(x) h(x) \\ &\times \left[ \ln \beta_i - \ln 2 - \ln \alpha_i - \ln \Gamma(1/\beta_i) \right. \\ &\left. - \alpha_i^{-\beta_i} |x - \mu_i|^{\beta_i} \right]. \end{aligned} \quad (8)$$

The quantity  $z_i(x)$  can be estimated as the conditional expectation of  $z_i$  given the observation  $x$  and the set of parameters  $\theta$  [30].

The EM algorithm consists in expectation and maximization steps, which are iterated until convergence. The expectation step is represented by the computations of the  $z_i(x)$  ( $i = 1, 2$  and  $x \in [0, L - 1]$ ) by using the current estimates of the parameters  $P_i$  and  $\theta_i$  ( $i = 1, 2$ ). The maximization step allows updating such parameter estimates. It is possible to show that the equations of these steps are as follows (see Appendix A):

- E-step: Compute  $z_i(x)$  at iteration  $t$  given the parameter estimates from the previous M-step:

$$z_i^{(t)}(x) = \frac{P_1^{(t)} p_1(x|\theta_1^{(t)})}{P_1^{(t)} p_1(x|\theta_1^{(t)}) + P_2^{(t)} p_2(x|\theta_2^{(t)})}. \quad (9)$$

- M-step: Maximize (8) given  $z_i^{(t)}(x)$  and  $\theta_i^{(t)}$ , which is equivalent to updating the statistical parameters as follows:

$$P_i^{(t+1)} = \sum_{x=0}^{L-1} z_i^{(t)}(x) h(x), \quad (10)$$

$$\sum_{x=0}^{L-1} z_i^{(t)}(x) h(x) \eta(x) |x - \mu_i^{(t+1)}|^{\beta_i^{(t)} - 1} = 0, \quad (11)$$

$$\alpha_i^{(t+1)} = \left[ \frac{\beta_i^{(t)} \sum_{x=0}^{L-1} z_i^{(t)}(x) h(x) |x - \mu_i^{(t)}|^{\beta_i^{(t)}}}{\sum_{x=0}^{L-1} z_i^{(t)}(x) h(x)} \right]^{1/\beta_i^{(t)}}, \quad (12)$$

$$\begin{aligned} \sum_{x=0}^{L-1} z_i^{(t)}(x) h(x) \left[ \frac{1}{\beta_i^{(t+1)}} + \frac{\Psi(1/\beta_i^{(t+1)})}{(\beta_i^{(t+1)})^2} \right. \\ \left. - \left( \frac{|x - \mu_i^{(t)}|}{\alpha_i^{(t)}} \right)^{\beta_i^{(t+1)}} \ln \left( \frac{|x - \mu_i^{(t)}|}{\alpha_i^{(t)}} \right) \right] = 0, \end{aligned} \quad (13)$$

where  $\eta(x)$  is given by

$$\eta(x) = \begin{cases} 1 & \text{if } x - \mu_i^{(t+1)} < 0, \\ -1 & \text{if } x - \mu_i^{(t+1)} \geq 0 \end{cases} \quad (14)$$

and  $\Psi(\cdot)$  is the digamma function defined as:  $\Psi(v) = \Gamma'(v)/\Gamma(v)$  [31].

According to the EM algorithm, the final parameter estimates are obtained starting from a set of initial values  $\theta^{(0)} = \{P_i^{(0)}, \mu_i^{(0)}, \alpha_i^{(0)}, \beta_i^{(0)} : i = 1, 2\}$  and then iterating the above equations until convergence. Eqs. (11) and (13) related to the estimation of the mean and shape parameters of the two class distributions, respectively, are nonlinear. Accordingly, they should be solved using numerical procedures such as the Newton–Raphson method [32]. It is worth noting that, while the convergence of the algorithm to a local maximum of the log likelihood is guaranteed, the identification of the global maximum cannot be assured. Indeed, the goodness of the solution found by the EM algorithm strongly depends on the choice of the initial parameter vector  $\theta^{(0)}$ . In order to cope with this issue, we propose a robust initialization strategy based on GAs, which will be described in Section 4.

### 3.3. Binary map generation

Without the need to compute the optimum threshold value explicitly, the adopted formulation of the problem allows to generate the thresholded map by utilizing any Bayesian decision criterion, such as the minimum error and minimum cost criteria. If we adopt the minimum error criterion to generate the binary map, since the final estimates of  $z_i(x)$  (with  $i = 1, 2$ ) represent the estimates of the posterior probabilities of the two classes (object and background), one can assign the optimal label  $y(m, n)$  to each pixel  $x(m, n)$  of the image  $X$  in such a way that

$$y(m, n) = \begin{cases} 0 & \text{if } z_1[x(m, n)] \geq z_2[x(m, n)], \\ 255 & \text{otherwise.} \end{cases} \quad (15)$$

If one wishes to compute explicitly the optimal threshold value  $T^*$  according to the MAP rule, a search must be made for the intersection point of the two posterior probability

functions  $z_1(x)$  and  $z_2(x)$ , i.e.,

$$z_1(T^*) = z_2(T^*). \tag{16}$$

#### 4. Initialization procedure for the EM algorithm

As mentioned in Section 3, the EM algorithm is sensitive to the problem of choosing the initial values of the parameters to be estimated. If these initial values are inappropriately selected, the EM algorithm may lead to an unsatisfactory estimation of the class distributions. To address this issue, several methods are reported in the literature. Some of these are based on the use of multiple random initial conditions to generate multiple solutions and then chose the one that produces the highest likelihood [33,34]. Others are based on initialization by clustering algorithms [33,35] or under a tree structure scheme [27].

In this paper, we propose to use GAs as an alternative solution to the problem of initialization. GAs represent a well-known family of methods for global optimization that have proved very attractive for their robustness explained by the facts that: (i) at each iteration of the optimization process, they retain a large number of candidate solutions and (ii) they do not base the search on the gradient principle but directly on the function to optimize [36,37]. The choice to adopt GAs for initializing the parameters of the distributions of classes mainly depends on the aforementioned properties of this optimization method. Since the initial values of the statistical parameters of classes strongly affect both the estimates obtained by the EM algorithm at convergence and the thresholding results, it is important to adopt an accurate and robust initialization procedure, which can properly explore the space of solutions. In GAs the effectiveness in the exploration of the space of the solutions is merged with the possibility to define an evaluation criterion based on a fitness function that implements the concepts adopted both by the EM algorithm (maximization of the log-likelihood function of the estimates) and by the threshold selection algorithm (minimization of the estimated error probability).

GAs perform a search by evolving a population of candidate individuals modeled with “chromosomes”. From one generation to the next, the population (set of candidate individuals) is improved by mechanisms inspired from genetics, i.e., through the use of both deterministic and non-deterministic genetic operators. The simplest form of a GA involves the following steps. First, an initial population of chromosomes (individuals) is generated randomly. Second, the goodness of the chromosomes is evaluated according to a predefined fitness function, which allows keeping the best chromosomes after rejecting the worst (the better the fitness, the higher the chance of being selected). This selection process is important for the next step, which is dedicated to reproducing the population. This is performed by means of genetic operators, such as crossover and mutation. The crossover operator crosses couples of individuals

(parents) with a probability  $P_c$  to form new offspring (children). If no crossover is performed for a given couple, the offspring is the exact copy of the parents. In order to introduce some randomness in the search process, the mutation operator is used. This latter mutates with a probability  $P_M$  each locus (i.e., each position in the chromosome) of each new offspring. This process is iterated until a user-defined convergence criterion is reached. The proposed GA-based search procedure adapted to our problem of identifying the optimal set of parameter values for initializing the EM algorithm is described in detail in the following.

Let us consider a population of  $S$  chromosomes  $C_m$  ( $m = 1, 2, \dots, S$ ) generated randomly. The chromosome  $C_m \in \mathfrak{R}^8$  is viewed as a vector representing the statistical parameters defining the mixture in (2) i.e.,  $C_m = [P_{1m}, P_{2m}, \mu_{1m}, \mu_{2m}, \alpha_{1m}, \alpha_{2m}, \beta_{1m}, \beta_{2m}]$ . Let  $f(m)$  be the fitness function value associated with the  $m$ th chromosome  $C_m$ . We propose to define  $f(m)$  as a function of the average between the normalized absolute log-likelihood value  $|L_m^*(X, Z, \theta^*)|_N$  and the estimated probability of error  $P_{em}^*$  obtained at the convergence of the EM algorithm by adopting the chromosome  $C_m$  for initialization, i.e.,

$$f(m) = e^{-0.5(|L_m^*(X, Z, \theta_m^*)|_N + P_{em}^*)}, \tag{17}$$

where the normalized absolute log likelihood is given by

$$|L_m^*(X, Z, \theta^*)|_N = \frac{|L_m^*(X, Z, \theta^*)|}{\sum_{i=1}^S |L_i^*(X, Z, \theta^*)|} \tag{18}$$

and the probability of error is computed according to the following equation:

$$P_{em}^* = \int_{x>T} P_{1m}^* p_{1m}(x|\theta_{1m}^*) dx + \int_{x\leq T} P_{2m}^* p_{2m}(x|\theta_{2m}^*) dx. \tag{19}$$

The use of this empirical expression for the fitness function is motivated by the fact that it is aimed at achieving two main goals: (i) to obtain an accurate fitting of the mixture of the estimated density functions to the observable global density function (i.e.,  $h(x)$ ), which is equivalent to minimizing the normalized absolute value of the log likelihood function over the mixture parameters and (ii) to yield a binary map with the minimum error, which corresponds to minimizing the estimated error probability  $P_{em}^*$ . The average between these two values is a reasonable empirical choice to satisfy both constraints. The absolute value of the log likelihood  $|L_m^*(X, Z, \theta_m^*)|$  is used instead of  $L_m^*(X, Z, \theta_m^*)$ , since the latter is always negative. The minimization of both the probability of error and the normalized absolute value of the log likelihood leads to the maximization of the fitness function (17). Accordingly, the higher the fitness value of a given chromosome, the higher the chance this latter has to be selected.

The general description of the proposed GA-based initialization procedure is summarized in Fig. 2. The algorithm

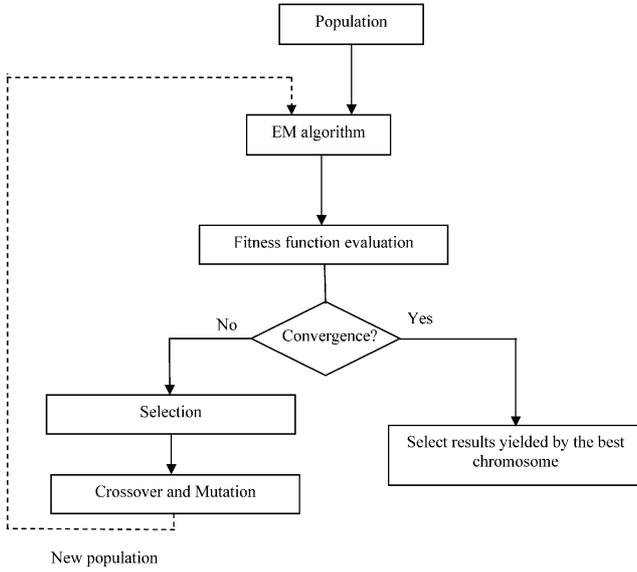


Fig. 2. Flow chart of the proposed GA-based initialization procedure.

starts by randomly generating an initial population of  $S$  chromosomes. The EM algorithm is then applied to each chromosome in order to estimate the statistical parameters of the object and background classes. The fitness function is evaluated for each chromosome after convergence of the EM algorithm according to (17). The chromosomes are then subject to selection, crossover, and mutation operators in order to generate a new population of size  $S$  as described in the following.

(1) *Selection operation*: The chromosomes in the current population are selected for reproduction according to the roulette wheel spinning method [36]. This method starts by assigning a probability  $q_m$  and a cumulative probability  $Q_m$  to each chromosome  $C_m$  such that

$$q_m = \frac{f(m)}{\sum_{j=1}^S f(j)}, \quad m = 1, \dots, S, \quad (20)$$

$$Q_m = \sum_{j=1}^m q_j, \quad m = 1, \dots, S. \quad (21)$$

Then a random nonzero floating point number  $l \in (0, 1]$  is generated. The chromosome  $C_m$  is chosen if  $Q_{m-1} < l \leq Q_m$ , ( $Q_0 = 0$ ). With this selection process, the higher the fitness function value of a chromosome, the higher the chance of being selected more times. At the next iteration, another chromosome is selected with a new random value of  $l \in (0, 1]$ . The two chromosomes are thus put together to form a couple. The selection process is iterated until  $S/2$  chromosomes organized in  $S/4$  couples ready for crossover and mutation operations are selected.

(2) *Crossover operation*: Let  $C_{old1}$  and  $C_{old2}$  be two candidate chromosomes selected in the previous step as a couple to produce new offspring (i.e.,  $C_{new1}$  and  $C_{new2}$ ) by crossover. This is accomplished by first generating a ran-

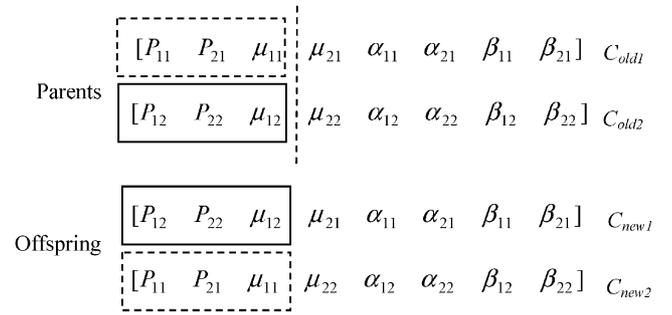


Fig. 3. Offspring generated by simple crossover of the two parent chromosomes.

dom value between  $[0, 1]$ . If this value is smaller than a user-predefined crossover probability  $P_C$ , the two candidate chromosomes undergo a crossover operation; otherwise, the new offspring ( $C_{new1}$  and  $C_{new2}$ ) are taken as the exact copy of their parents ( $C_{old1}$  and  $C_{old2}$ ). In this paper, two simple crossover operations are adopted. For each couple, one of these two operations is selected randomly. In the first operation, a crossover locus is chosen randomly and two offspring are generated by exchanging the values at that locus between the two parents (see Fig. 3). In the second operation, two offspring are generated according to an arithmetical operation between the two parent chromosomes. The first offspring is generated by averaging the values of the two chromosomes, i.e.,

$$C_{new1} = \frac{1}{2}(C_{old1} + C_{old2}) \quad (22)$$

while the second offspring is obtained by computing the inner root of the product (i.e.,  $\langle \cdot, \cdot \rangle^{1/2}$ ) between the two chromosomes, i.e.,

$$C_{new2}(j) = \langle C_{old1}, C_{old2} \rangle^{1/2}. \quad (23)$$

(3) *Mutation operation*: After performing the crossover operation for each couple  $C_{old1}$  and  $C_{old2}$ , the two new corresponding offspring (i.e.,  $C_{new1}$  and  $C_{new2}$ ) are subject to a mutation operation. This prevents the search process from falling into a local optimum solution. The mutation operation is applied to an offspring if a randomly generated value between  $[0, 1]$  is smaller than the user-defined probability of mutation  $P_M$ . Then, a position in the offspring is selected randomly and the corresponding value is perturbed by a random noise following a uniform distribution [38]. At the end of the mutation operation, a new generation of the  $S/2$  chromosomes selected from the previous generation and their corresponding  $S/2$  offspring is obtained.

(4) *Convergence criterion*: The entire process made up of the EM, the selection, the crossover, and the mutation steps is repeated until a stop criterion is satisfied. A simple stop criterion consists of evaluating the absolute difference between the average values of fitness functions  $F_{av}$  computed over two successive generations (i.e., two successive

iterations  $r - 1$  and  $r$ ).

$$|F_{av}(r) - F_{av}(r - 1)| < \delta, \quad (24)$$

where

$$F_{av}(j) = \sum_{m=1}^S \frac{f^j(m)}{S}, \quad j = r - 1, r \quad (25)$$

and  $\delta$  is an arbitrary small positive constant.

## 5. Experimental results

In order to assess the effectiveness of the proposed thresholding method, both simulated and real images were used in the experiments. In particular, simulated data with different statistical characteristics were considered as well as real images related to nondestructive test (NDT) and remote sensing (RS) problems. The results provided by the proposed method were compared with those yielded by a set of four thresholding techniques widely used in the literature, i.e. Otsu's [11], K&I's [14], Kapur's [13] and H&W's [20] techniques. In all experiments, we adopted the following parameters for the GA-based initialization procedure: population size  $S = 50$ , crossover probability  $P_C = 0.9$ , mutation probability  $P_M = 0.1$ , and convergence threshold  $\delta = 0.001$ .

### 5.1. Experiments on simulated data

In the experiments on simulated data, we analyzed the robustness of the proposed method to three important features that characterize a histogram composed of a mixture of two distributions: (1) the prior probability of each distribution; (2) their degree of overlap; and (3) the statistical model they follow. Three experiments were designed to test the robustness of the method to these three critical issues. For each experiment, several simulated test images with histograms composed of mixtures with different statistical characteristics associated with the object and background classes were generated (see Fig. 4 for an example of these images).

In the first experiment, in order to evaluate the effect of the prior probability parameter on the proposed method, four test images representing Gaussian mixtures were produced by varying the prior probabilities and keeping the means (i.e.,  $\mu_1 = 50$ ,  $\mu_2 = 100$ ), and standard deviations constant (i.e.,  $\sigma_1 = \sigma_2 = 15$ ). As an example, Fig. 5 shows the histogram of a test image with a mixture characterized by equal prior probabilities (i.e.,  $P_1 = P_2 = 0.5$ ). The threshold values obtained by applying the different thresholding methods to the four test cases are reported in Table 1 (in this table the optimal threshold value corresponds to the optimal Bayes threshold, which is obtained from the intersection between the two true class densities weighted by their true prior probabilities). As can be seen, in all test cases the thresholds obtained by the proposed method are the optimal

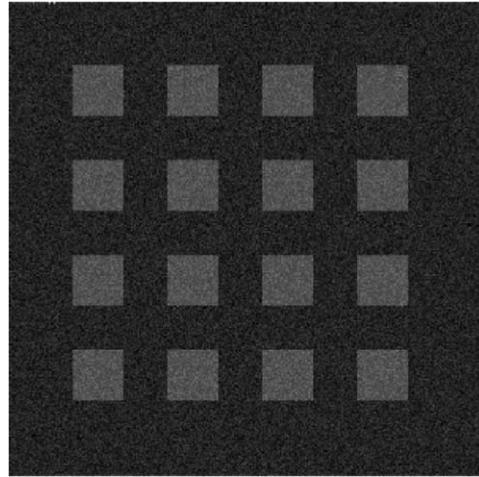


Fig. 4. Example of a test image used in the experiments simulating object and background.

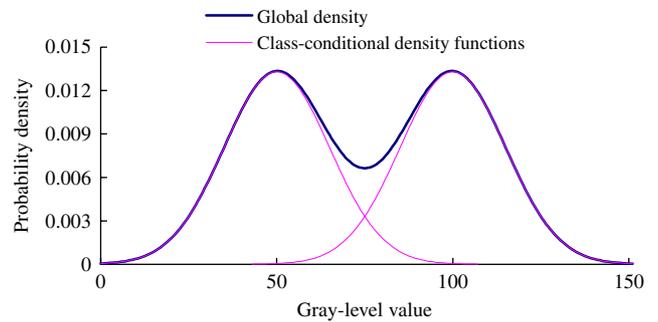


Fig. 5. Test histogram generated by simulating a Gaussian mixture with equal prior probabilities  $P_1 = P_2 = 0.5$  ( $\mu_1 = 50$ ,  $\mu_2 = 100$ ,  $\sigma_1 = \sigma_2 = 15$ ).

ones. Among the reference thresholding algorithms, as theoretically expected in such situations [14], K&I's method proved almost as accurate as the proposed method since it provided threshold values very close to the optimal ones. The Kapur's algorithm seems dealing satisfactorily with the issue of unbalanced priors. By contrast, the two other non-parametric algorithms, namely Otsu's and H&W's methods, performed well only when the population sizes of the two classes were similar (i.e., in situations where prior probabilities were balanced) confirming the experimental results found in Refs. [17,23–26].

As previously mentioned, the second experiment aimed at assessing the sensitivity of the proposed method to the degree of overlap between the distributions of the object and background classes. Also in this case, different test images were generated to simulate Gaussian mixtures characterized by fixed prior probabilities and standard deviations (i.e.,  $\sigma_1 = \sigma_2 = 15$ ) but different class means,  $\mu_1$  and  $\mu_2$ , to vary the degree of overlap. Regarding the choice of the prior probability values, we considered two different scenarios: balanced and unbalanced prior probabilities (i.e.,  $P_1 = P_2 = 0.5$ , and  $P_1 = 0.9$ ,  $P_2 = 0.1$ , respectively). As an

Table 1

Threshold values obtained on simulated histograms, generated to test the sensitivity of the different thresholding techniques to the variation of the prior probabilities

Test case	$(P_1, P_2)$	$T_{\text{optimal}}$	$T_{\text{EM-GA}}$	$T_{\text{K\&I}}$	$T_{\text{Otsu}}$	$T_{\text{H\&W}}$	$T_{\text{Kapur}}$
1	(0.9, 0.1)	85	85	89	64	53	78
2	(0.8, 0.2)	82	82	84	71	72	74
3	(0.6, 0.4)	77	77	77	74	74	73
4	(0.5, 0.5)	75	75	74	75	73	75
Average time		—	114.87	1.796	1.750	1.765	1.703

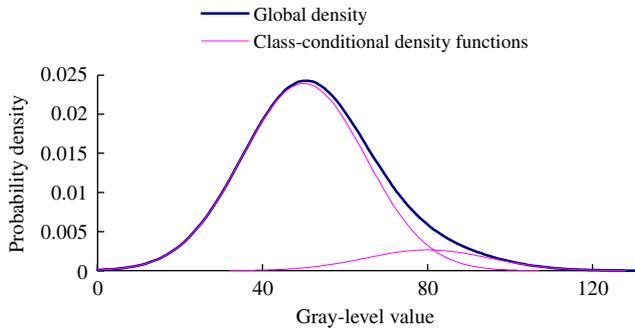


Fig. 6. Test histogram produced with a Gaussian mixture characterized by the parameters  $P_1 = 0.9$ ,  $P_2 = 0.1$ ,  $\mu_1 = 50$ ,  $\mu_2 = 80$ ,  $\sigma_1 = \sigma_2 = 15$ .

example of the second scenario, Fig. 6 depicts the histogram of a test image representing a strongly overlapped mixture, where the means of the two classes are 50 and 70, respectively. For both scenarios, Table 2 shows the threshold values obtained by the different methods on the different simulated images. In the first scenario, where the priors are balanced, the results point out that the five methods perform very well up to reasonable overlapped mixtures (i.e., test cases 1–4 in Table 2a). When the overlap becomes strong (i.e., test cases 5, 6 in Table 2a), both the proposed method and K&I's faced problems since they could not provide a meaningful threshold value, while the three other reference methods could recover the optimal threshold. More specifically, in the test cases 5, 6, the global histogram representing the mixtures became monomodal and nearly Gaussian. Consequently, at convergence of the EM–GA algorithm, the estimated mixture corresponded to two nearly Gaussian distributions: one with prior probability almost equal to one and centered at the optimal threshold value, the other with negligible prior probability. The K&I criterion did not exhibit any internal minimum but a clear maximum at the optimal threshold value, specifying that the maximum error is reached at that point. These results are justified by the fact that the minimum error criterion, exploited by these two methods in their formulation, is applicable only if the global histogram is multimodal. By contrast, the other three reference methods have a tendency to split the only mode in the middle because of the properties of the cost function they use. In the second scenario, we again considered six test cases, this time by choosing unbalanced prior probability values ( $P_1 = 0.9$ ,

$P_2 = 0.1$ ). In all the test cases (Table 2b), the best accuracy in the threshold estimation was obtained by the proposed method, which proved very stable also in the critical case of very overlapped distributions (i.e.,  $\mu_1 = 50$ ,  $\mu_2 = 70$ ). In greater detail, with the proposed technique it was possible to identify the optimal threshold in the first three test cases. However, for difficult situations characterized by very overlapped distributions in addition to unbalanced priors, an increase in the error between the optimal and the computed thresholds can be observed. Such an increase, anyway is still limited. As an example, the optimal threshold and the one computed by the proposed method are equal to 82 and 84 for  $(\mu_1, \mu_2) = (50, 80)$ , and become equal to 85 and 90, respectively, for  $(\mu_1, \mu_2) = (50, 70)$ . These experiments show that the other methods are less stable than the proposed one.

It is worth noting that, according to the results obtained in the two above experiments, which were carried out under the Gaussian assumption for the object and background classes, the proposed method proved to be more effective than the K&I method, whose parametric formulation is based on the Gaussian assumption. This can be explained by the fact that: (1) the GG model is also capable of approximating a Gaussian distribution accurately and (2) contrary to the K&I method, the EM-based estimation procedure together with the GA-based initialization strategy represents an effective way of obtaining unbiased estimates of the statistical parameters of the object and background classes, even in unfavorable conditions of prior probabilities and overlap between the two classes.

In the third experiment, we assessed the full capabilities of the proposed method, i.e., by considering non-Gaussian statistical models for the object and background classes. To carry out such a task, we generated four test images by simulating mixtures of GG distributions characterized by different shape values. The remaining statistical parameters were kept unchanged and set as follows:  $P_1 = 0.9$ ,  $P_2 = 0.1$ ,  $\mu_1 = 50$ ,  $\mu_2 = 100$ ,  $\sigma_1 = \sigma_2 = 15$ . Fig. 7 shows an example of the histogram of a test image with shape parameters  $\beta_1$  and  $\beta_2$  equal to 1 and 6 corresponding to Laplacian and pseudouniform distributions, respectively. Table 3 shows the thresholds obtained by the different methods on the four test cases. Again in this experiment, the results point out that the proposed method is capable of providing very accurate thresholds, since for all four test cases, the obtained

Table 2

Threshold values obtained on simulated histograms generated to test the sensitivity of the different thresholding techniques to the overlap between object and background class distributions with: (a) equal probabilities ( $P_1 = P_2 = 0.5$ ) and (b) unbalanced prior probabilities ( $P_1 = 0.9, P_2 = 0.1$ )

Test case	$(\mu_1, \mu_2)$	$T_{\text{Optimal}}$	$T_{\text{EM-GA}}$	$T_{\text{K\&I}}$	$T_{\text{Otsu}}$	$T_{\text{H\&W}}$	$T_{\text{Kapur}}$
<i>(a)</i>							
1	(50, 120)	85	85	84	85	85	85
2	(50, 110)	80	80	79	80	80	80
3	(50, 100)	75	75	74	75	73	75
4	(50, 90)	70	70	69	70	70	70
5	(50, 80)	65	—	—	65	65	65
6	(50, 70)	60	—	—	60	60	60
Average time		—	72.53	1.796	1.750	1.765	1.703
<i>(b)</i>							
1	(50, 120)	92	92	93	83	87	79
2	(50, 110)	88	88	90	75	51	79
3	(50, 100)	85	85	89	64	53	78
4	(50, 90)	83	82	92	58	53	78
5	(50, 80)	82	84	107	55	52	77
6	(50, 70)	85	90	121	52	51	72
Average time		—	99.18	1.796	1.750	1.765	1.703

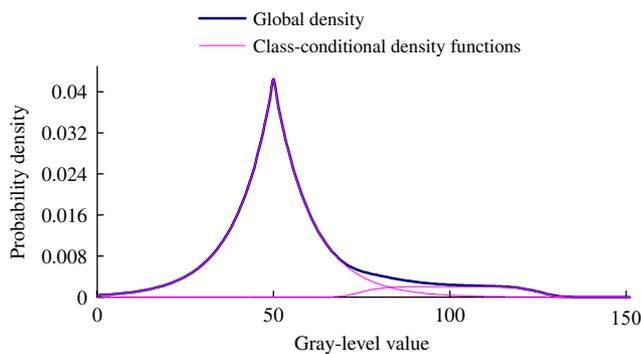


Fig. 7. Test histogram generated by simulating a GG mixture with parameter values  $P_1 = 0.9, P_2 = 0.1, \mu_1 = 50, \mu_2 = 100, \sigma_1 = \sigma_2 = 15, \beta_1 = 1, \beta_2 = 6$ .

threshold values fit optimal ones exactly. Among the non-parametric methods used for comparison (i.e., Kapur's, H&W's, and Otsu's algorithms), Kapur's algorithm proved to be the best, but it gave less accurate results than those of the proposed method, in particular, in cases where the mixture includes a Laplacian distribution. This can be explained by the fact that the entropy of small value shape parameter distributions such as the Laplacian is low, thus forcing Kapur's algorithm (which is based on the maximum entropy principle) to select biased threshold solutions for which the entropy-based criterion is higher. Due to its criterion being limited to first-order statistics, H&W's algorithm was not able to discriminate between the complex distributions adopted in these experiments. The poor performance of Otsu's algorithm can be found in the unbalanced prior probabilities used to construct the GG mixture. The parametric K&I's algorithm could not run as well as in the previous experiments because of the Gaussian assumption. Although the superiority of the proposed method was

expected because the mixtures are based on GG distributions, the obtained promising results allowed also to validate our GA-based initialization procedure for the parameter estimation process.

## 5.2. Experiments on NDT images

The second set of experiments is related to the problem of the analysis of NDT images. NDT consists of the use of special equipments and methods to detect an object and quantify its possible defects without harming it. NDT methods are used in a broad variety of applications, such as nuclear industry, chemistry, aeronautics and astronautics, civil constructions, etc. In this work, two real NDT images were used in the experiments to assess the performance of our method. The first image (of size  $256 \times 256$  pixels) represents a light microscopy image of a material structure. Light microscopy is frequently used for inspecting the microstructures of materials in order to derive information about their properties such as porosity, particle sizes, distribution uniformity, etc. The second NDT image (of size  $256 \times 128$  pixels) depicts a defective cloth. The two images (with their corresponding ground truth) are shown in Figs. 8 and 9, respectively.

The results in terms of threshold values and of number of misclassified pixels obtained by applying each thresholding method to the first NDT image (whose histogram is reported in Fig. 10a) are reported in Table 4a (it is worth noting that all the pixels were assigned to one of the two classes without any rejection option). They show that the threshold value yielded by the proposed method was the closest to the optimal one. In this case, it was equal to 184, corresponding to 1151 misclassified pixels, while the optimal threshold (found by manually thresholding the image on the basis of the information present in the ground-truth

Table 3

Threshold values obtained on simulated histograms generated to test the sensitivity of the different thresholding techniques to the shapes of the distribution of the object and background classes

Test case	$(\beta_1, \beta_2)$	$T_{\text{optimal}}$	$T_{\text{EM-GA}}$	$T_{\text{K\&I}}$	$T_{\text{Otsu}}$	$T_{\text{H\&W}}$	$T_{\text{Kapur}}$
1	(1, 1)	87	87	82	70	77	73
2	(1, 6)	83	83	85	69	71	74
3	(6, 2)	81	81	91	58	49	79
4	(6, 6)	81	81	94	58	49	79
Average time		—	110.54	1.796	1.750	1.765	1.703

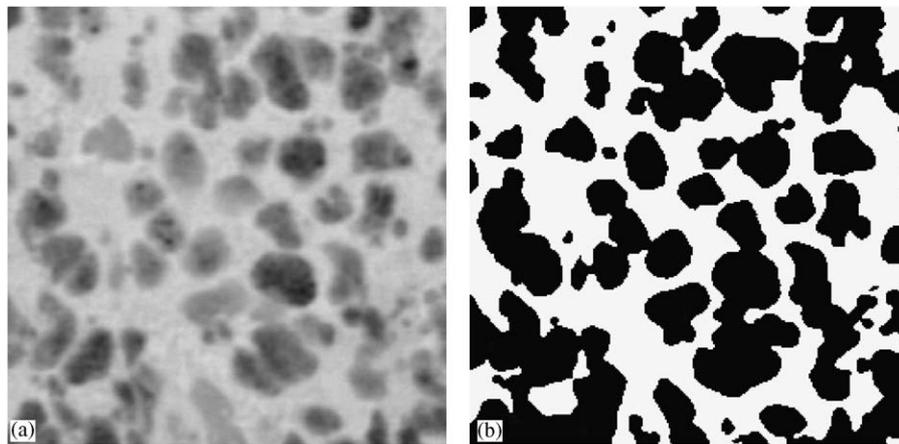


Fig. 8. First NDT problem: (a) material image and (b) ground-truth image.

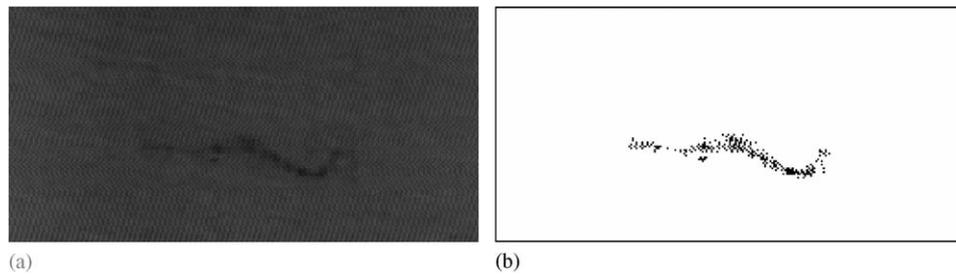


Fig. 9. Second NDT problem: (a) defective cloth image and (b) ground-truth image.

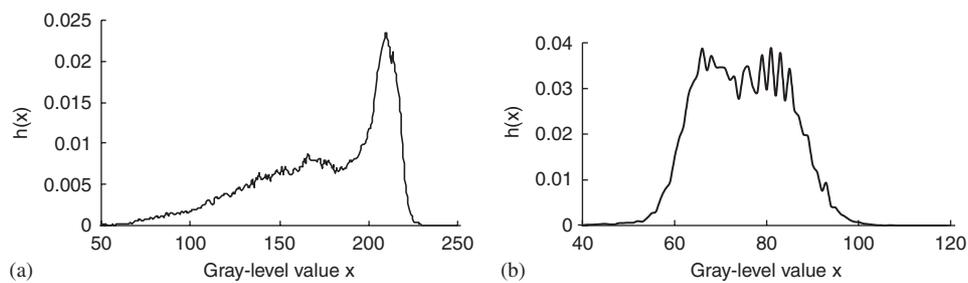


Fig. 10. Histograms of the two NDT images: (a) material image and (b) defective cloth image.

Table 4

Threshold values and numbers of misclassified pixels obtained with the different thresholding methods on: (a) the NDT material image and (b) the NDT defective cloth image

	Optimal	EM–GA	K&I	Otsu	H&W	Kapur
<i>(a)</i>						
Threshold value	186	184	193	166	181	136
Misclassified pixels	551	1151	3613	9836	2342	22742
Time (ms)	—	76.953	1.625	1.625	1.656	1.609
<i>(b)</i>						
Threshold value	55	54	36	74	74	53
Misclassified pixels	90	92	243	15 844	15 844	97
Time (s)	—	33.437	1.937	1.921	2	1.875

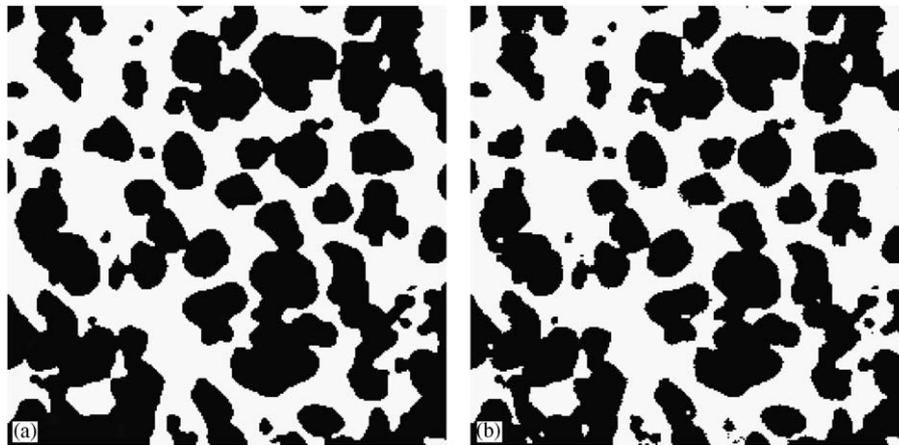


Fig. 11. NDT material binary images obtained by: (a) manual supervised optimal thresholding and (b) the proposed thresholding method.

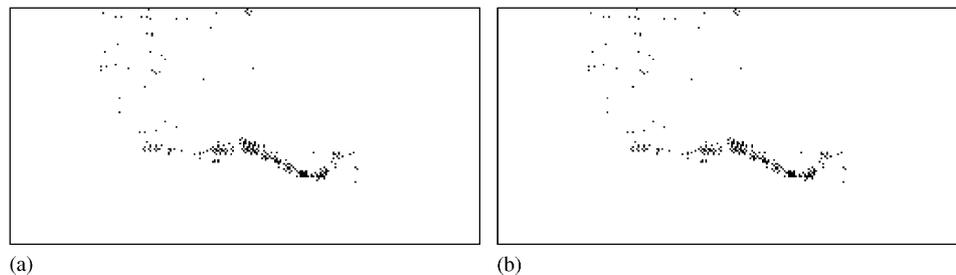


Fig. 12. NDT defective cloth binary images obtained by: (a) manual supervised optimal thresholding and (b) the proposed thresholding method.

image) was equal to 186, corresponding to 551 misclassified pixels (see Fig. 11 for a visual comparison between the binary image obtained by the proposed method and the one yielded by manual optimal thresholding). Among the reference thresholding techniques, the worst result was provided by Kapur's method, which totally failed to determine a reasonable threshold, while the threshold closest to the one yielded by the proposed method was computed by H&W's algorithm.

The results obtained by thresholding the histogram (see Fig. 10b) of the cloth image are provided in Table 4b. Also in this case, the best threshold value was given by the

proposed method. It was equal to 54, corresponding to 92 misclassified pixels, while the optimal threshold obtained by manual supervised thresholding was equal to 55, corresponding to 90 misclassified pixels. This promising result is confirmed by visually inspecting the binary images obtained both by the proposed method and by manual optimal thresholding (see Fig. 12). Similarly to what was observed in the experiments with data simulating a strong overlap between the distributions of the two classes (see previous Section 5.1), Kapur's method performed well for this real image (the threshold value was equal to 53, corresponding to 97 misclassified pixels), while the worst result was obtained by

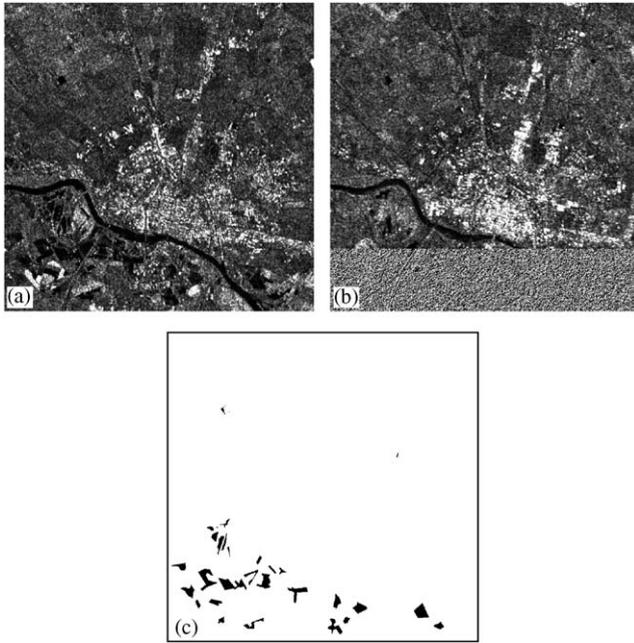


Fig. 13. SAR images related to the city of Pavia (Italy): (a) image acquired on 20th October 2000 immediately after flooding; (b) image acquired on 28th October 2000 when the flooded areas were reduced; and (c) map of changed areas used as ground truth.

Otsu's and H&W's methods (their threshold value was equal to 74, corresponding to 15 844 misclassified pixels).

### 5.3. Experiments on multitemporal SAR images

In these experiments, we propose to apply our method to the complex problem of the detection of changed (and unchanged) areas on the Earth's surface by the analysis of multitemporal images acquired by a spaceborne Synthetic Aperture Radar (SAR). In particular, two SAR images acquired by the ERS-2 sensor over the city of Pavia (Italy) on

20 and 28 October 2000, respectively, were considered. The two images represent a flooding event that occurred during the two acquisition dates. The equivalent numbers of looks (ENL) of the two images are equal to 2.64 and 3.43, respectively. These values mean that the two images are highly corrupted by a critical multiplicative speckle noise, which is typical of SAR data. The considered images (with the corresponding ground-truth image) are shown in Fig. 13.

In order to reduce the speckle noise effects that make separation between changed and unchanged classes difficult, we applied two iterations of the enhanced Lee filter with a window size of  $5 \times 5$  pixels to the two original SAR images [39]. Then, according to what usually done in the context of change detection in SAR imagery [40], changes can be identified by thresholding the log-ratio image generated from the filtered images (see Fig. 14). Table 5 shows the results achieved with the different thresholding methods. The threshold value obtained by the proposed method was the closest to the optimal threshold value provided by manual supervised thresholding of the log-ratio image. In this case, the threshold obtained by the proposed method was equal to 110 (corresponding to 5692 misclassified pixels), while the optimal threshold was equal to 106 (corresponding to 5536 misclassified pixels). Concerning the reference thresholding techniques, on the one hand, H&W's and Otsu's methods failed to obtain a reasonable threshold. This can be explained by the small number of pixels representing the changed class (i.e., by the unbalanced prior probabilities of the object and background classes), which penalizes the aforementioned methods, as already shown in the experiments with simulated data (see Section 5.1). On the other hand, the K&I's and the Kapur's methods exhibited in this case results competing with the proposed method. A visual comparison between the change-detection map provided by the proposed method and that yielded by manual supervised thresholding confirms the effectiveness of the presented threshold-selection method (see Fig. 15).

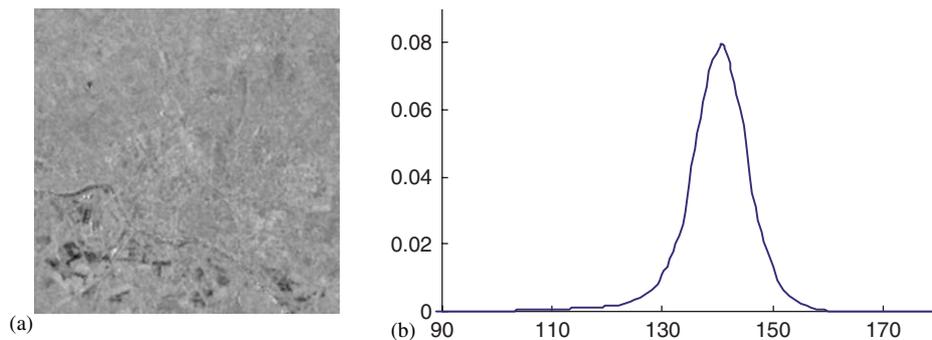


Fig. 14. (a) Log-ratio image generated from filtered images and (b) corresponding histogram.

Table 5

Threshold values and numbers of misclassified pixels obtained with the different thresholding methods on the log-ratio SAR image of the city of Pavia (Italy)

	Optimal	EM-GA	K&I	Otsu	H&W	Kapur
Threshold value	106	110	117	152	147	116
Misclassified pixels	5536	5692	7138	268 459	171 622	6754
Time (ms)	—	91.094	1.515	1.531	1.531	1.515

## 6. Conclusions

In this paper, a novel parametric image thresholding method has been presented. The underlying idea of the proposed method is to estimate the statistical parameters related to object and background classes by using the EM algorithm under the assumption that the two classes follow a GG distribution. The main contributions of this paper are related to two important methodological issues. The first one is related to the formulation of the EM estimation algorithm under a GG-based mixture model. The choice of this distribution is attractive for modeling the object and background classes in a generic image under analysis, since it is flexible and robust. In other words, despite it can approximate a large class of statistical distributions, it requires only the estimation of one additional parameter (i.e., the shape parameter) compared to the widely used Gaussian model. The second contribution concerns the critical problem of parameter initialization in the EM algorithm. This has been addressed by developing a robust and efficient strategy based on GAs.

In order to assess the effectiveness of the proposed method, in the experiments we considered several different simulated and real images. For purposes of comparison, we also used four thresholding methods widely referenced in the literature. From the obtained results, it was possible to underline the following main attractive properties of the proposed method: (i) it showed very high thresholding accuracies (in average it proved to be the most effective among the considered techniques and provided results that were very close to those yielded by optimal manual supervised thresholding); (ii) it can address situations characterized by strongly unbalanced prior probabilities between object and background; and (iii) it is robust to the problem of ambiguity between object and background classes in non-homogeneous images. The only critical issue we verified in our experiments with the proposed technique concerns cases in which the histogram of the considered image is monomodal, due both to very overlapped class distributions and to balanced prior probabilities. In these extreme critical situations, traditional techniques proved to be more effective than the proposed method. Another advantage of the proposed method lies in the possibility of applying any kind of decision rule (e.g., minimum cost, minimum risk, spatial context-based decision rules [41]) for selecting the optimal threshold value thanks to its explicit unbiased estimation of the class statistical parameters.

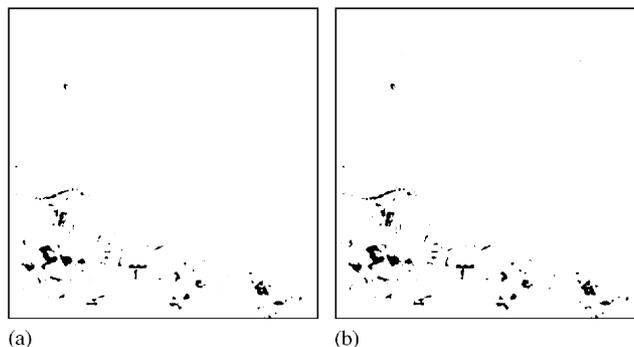


Fig. 15. Change-detection maps obtained by thresholding the log-ratio image with: (a) manual supervised optimal thresholding and (b) the proposed thresholding method.

The main disadvantage of the proposed technique consists in the higher computational time required compared to standard algorithms (which depends on the GA used for the identification of the best initial conditions). Nonetheless, this computational time is acceptable in many applications (on average in all our experiments the processing time was not greater than 2 min on a standard PC). In addition, it is worth noting that the computational complexity of the GA can be reduced either by using approximated strategies for the evaluation of the fitness function or by reducing the number of iterations required for reaching the convergence. This results in different trade-offs between required computational time and quality of the solution.

As a future development of this work, the proposed method will be extended to the problem of multilevel thresholding and will be integrated with a Markov random field approach for modeling spatial-context information in the thresholding process.

## Acknowledgment

The authors wish to thank Dr. Mehmet Sezgin (Tübitak Marmara Research Center, Turkey) and Dr. Paolo Gamba (University of Pavia, Italy) for providing the NDT and the multitemporal SAR images used in the experiments.

## Appendix A

In this appendix, we derive the EM equations for a mixture of two GG distributions. The complete data log likelihood

is given by

$$L(X, Z, \theta) = \sum_{i=1}^2 \sum_{x=0}^{L-1} z_i(x)h(x) \ln[P_i p_i(x|\theta_i)], \quad (26)$$

where  $z_i(x)$  can be estimated as the conditional expectation of  $z_i$  given the observation  $x$  and the parameter set  $\theta = [P_1, P_2, \theta_1, \theta_2]$ . In the case of a GG distribution, the complete data log likelihood is given as follows:

$$L(X, Z, \theta) = \sum_{i=1}^2 \sum_{x=0}^{L-1} z_i(x)h(x) \times \ln \left[ P_i \frac{\beta_i}{2\alpha_i \Gamma(1/\beta_i)} e^{-[|x-\mu_i|/\alpha_i]^{\beta_i}} \right]. \quad (27)$$

Eq. (27) can be written after some transformations in the following compact form:

$$L(X, Z, \theta) = \sum_{i=1}^2 \sum_{x=0}^{L-1} z_i(x)h(x) \ln P_i + \sum_{i=1}^2 \sum_{x=0}^{L-1} z_i(x)h(x) \times \left[ \ln \beta_i - \ln 2 - \ln \alpha_i - \ln \Gamma(1/\beta_i) - \alpha_i^{-\beta_i} |x - \mu_i|^{\beta_i} \right]. \quad (28)$$

The estimation of the mean ( $\mu_1, \mu_2$ ), the scale ( $\alpha_1, \alpha_2$ ) and the shape ( $\beta_1, \beta_2$ ) parameters associated with object and background classes is carried out by maximizing the above function as discussed in the following.

### A.1. Estimation of the mean $\mu_i$

The estimation of the mean is obtained by computing the derivative of (28) with respect to  $\mu_i$ , i.e.,

$$\frac{dL(X, Z, \theta)}{d\mu_i} = 0. \quad (29)$$

This allows to obtain

$$\sum_{x=0}^{L-1} z_i(x)h(x) \left[ -\frac{\beta_i}{\alpha_i^{\beta_i}} \eta(x) |x - \mu_i|^{\beta_i-1} \right] = 0, \quad (30)$$

where

$$\eta(x) = \begin{cases} 1 & \text{if } x - \mu_i < 0, \\ -1 & \text{if } x - \mu_i \geq 0. \end{cases} \quad (31)$$

By taking the term  $-\beta_i/\alpha_i^{\beta_i}$  out of the summation, we find that the estimate of the mean is achieved by solving the following nonlinear equation:

$$g(\mu_i) = \sum_{x=0}^{L-1} z_i(x)h(x)\eta(x)|x - \mu_i|^{\beta_i-1} = 0. \quad (32)$$

### A.2. Estimation of the scale parameter $\alpha_i$

The estimation of the scale parameter is yielded by computing the derivative of (28) with respect to  $\alpha_i$ , i.e.,

$$\frac{dL(X, Z, \theta)}{d\alpha_i} = 0. \quad (33)$$

Accordingly, we obtain

$$\sum_{x=0}^{L-1} z_i(x)h(x) \left[ -\frac{1}{\alpha_i} + \beta_i \alpha_i^{-\beta_i-1} |x - \mu_i|^{\beta_i} \right] = 0. \quad (34)$$

After some simple transformations, we get the estimate of the scale parameter from the following equation:

$$\alpha_i = \left[ \frac{\beta_i \sum_{x=0}^{L-1} z_i(x)h(x) |x - \mu_i|^{\beta_i}}{\sum_{x=0}^{L-1} z_i(x)h(x)} \right]^{1/\beta_i}. \quad (35)$$

### A.3. Estimation of the shape parameter $\beta_i$

Also for the shape parameter, the estimation is yielded by computing the derivative of (28) with respect to  $\beta_i$ , i.e.,

$$\frac{dL(X, Z, \theta)}{d\beta_i} = 0 \quad (36)$$

which is equivalent to

$$\sum_{x=0}^{L-1} z_i(x)h(x) \left[ \frac{1}{\beta_i} + \frac{\Gamma'(1/\beta_i)}{\beta_i^2 \Gamma(1/\beta_i)} - \left( \frac{|x - \mu_i|}{\alpha_i} \right)^{\beta_i} \times \ln \left( \frac{|x - \mu_i|}{\alpha_i} \right) \right] = 0, \quad (37)$$

where

$$\frac{d}{d\beta_i} \ln(\Gamma(1/\beta_i)) = -\frac{\Gamma'(1/\beta_i)}{\beta_i^2 \Gamma(1/\beta_i)}. \quad (38)$$

This can be rewritten as

$$\sum_{x=0}^{L-1} z_i(x)h(x) \left[ \frac{1}{\beta_i} + \frac{1}{\beta_i^2} \Psi(1/\beta_i) - \left( \frac{|x - \mu_i|}{\alpha_i} \right)^{\beta_i} \times \ln \left( \frac{|x - \mu_i|}{\alpha_i} \right) \right] = 0, \quad (39)$$

where  $\Psi(v) = \Gamma'(v)/\Gamma(v)$  is the digamma function [31].

Finally, the estimate of the shape parameter is achieved by solving the following nonlinear equation:

$$\varphi(\beta_i) = \sum_{x=0}^{L-1} z_i(x)h(x) \left[ \frac{1}{\beta_i} + \frac{1}{\beta_i^2} \Psi(1/\beta_i) - \left( \frac{|x - \mu_i|}{\alpha_i} \right)^{\beta_i} \ln \left( \frac{|x - \mu_i|}{\alpha_i} \right) \right] = 0. \quad (40)$$

A.4. Numerical evaluation of the mean and the shape parameters

In order to compute the mean and the shape parameter values at the EM iteration  $t + 1$  from the nonlinear equations (32) and (40), we use the standard iterative Newton–Raphson method. At iteration  $(r + 1)$  of the latter, the estimation of the mean value is given by (for simplicity the subscripts  $t$  and  $t + 1$  referring to the EM iterations are omitted in the expressions)

$$\mu_i^{(r+1)} = \mu_i^{(r)} - \frac{g(u_i^{(r)})}{g'(\mu_i^{(r)})}, \tag{41}$$

where

$$g'(\mu_i^{(r)}) = (\beta_i - 1) \sum_{x=0}^{L-1} z_i(x)h(x)|x - \mu_i^{(r)}|^{\beta_i-2}. \tag{42}$$

In the same manner, the shape parameter at step  $(r + 1)$  is given by

$$\beta_i^{(r+1)} = \beta_i^{(r)} - \frac{\varphi(\beta_i^{(r)})}{\varphi'(\beta_i^{(r)})}, \tag{43}$$

where

$$\begin{aligned} \varphi'(\beta_i^{(r)}) = & \sum_{x=0}^{L-1} z_i(x)h(x) \left[ -\frac{1}{(\beta_i^{(r)})^2} - \frac{\Psi'(1/\beta_i^{(r)})}{(\beta_i^{(r)})^4} \right. \\ & - \frac{2}{(\beta_i^{(r)})^3} \Psi(1/\beta_i^{(r)}) - \left. \left( \frac{|x - \mu_i|}{\alpha_i} \right)^{\beta_i^{(r)}} \right. \\ & \times \left. \left[ \ln \left( \frac{|x - \mu_i|}{\alpha_i} \right) \right]^2 \right] \end{aligned} \tag{44}$$

in which the function  $\Psi'(v)$  is the trigamma function defined as:  $\Psi'(v) = d^2/dv^2 \ln \Gamma(v)$  [31].

References

[1] T. Sund, K. Eilertsen, An algorithm for fast adaptive binarization with applications in radiotherapy imaging, *IEEE Trans. Med. Imaging* 22 (2003) 22–28.  
 [2] Y. Solihin, C.G. Leedham, Integral ratio: a new class of global thresholding techniques for handwriting images, *IEEE Trans. Pattern Anal. Mach. Intell.* 21 (1999) 761–768.  
 [3] B. Bhanu, Automatic target recognition: state of the art survey, *IEEE Trans. Aerosp. Electron. Syst.* AES-22 (1986) 364–379.  
 [4] L. Bruzzone, D.F. Prieto, Automatic analysis of the difference image for unsupervised change detection, *IEEE Trans. Geosci. Remote Sensing* 38 (2000) 1171–1182.  
 [5] L. Bruzzone, D.F. Prieto, An adaptive and semiparametric and context-based approach to unsupervised change detection in multitemporal remote sensing images, *IEEE Trans. Image Process.* 11 (2002) 452–466.  
 [6] P.L. Rosin, E. Ioannidis, Evaluation of global image thresholding for change detection, *Pattern Recognition Lett.* 24 (2003) 2345–2356.  
 [7] F.H.Y. Chan, F.K. Lam, H. Zhu, Adaptive thresholding by variational method, *IEEE Trans. Image Process.* 7 (1998) 468–473.

[8] F. Liu, X. Song, Y. Luo, D. Hu, Adaptive thresholding based on variational background, *Electron. Lett.* 38 (2002) 1017–1018.  
 [9] T.W. Ridler, S. Calvard, Picture thresholding using an iterative selection method, *IEEE Trans. Syst. Man Cybern.* SMC-8 (1978) 630–632.  
 [10] H.J. Trussell, Comments on picture thresholding using an iterative selection method, *IEEE Trans. Syst. Man Cybern.* SMC-9 (1979) 311.  
 [11] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Syst. Man Cybern.* SMC-9 (1979) 62–66.  
 [12] W.H. Tsai, Moment preserving thresholding: a new approach, *Comput. Vision Graph. Image Process.* 29 (1985) 377–393.  
 [13] J.N. Kapur, P.K. Sahoo, A.K.C. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, *Comput. Vision Graph. Image Process.* 29 (1985) 273–285.  
 [14] J. Kittler, J. Illingworth, Minimum error thresholding, *Pattern Recognition* 19 (1986) 41–47.  
 [15] A.K.C. Wong, P.K. Sahoo, A grey-level threshold selection method based on maximum entropy principle, *IEEE Trans. Syst. Man Cybern.* 19 (1989) 866–871.  
 [16] A.S. Abutaleb, Automated thresholding of gray-level pictures using two-dimensional entropies, *Comput. Vision Graph. Image Process.* 7 (1989) 22–32.  
 [17] T. Kurita, N. Otsu, N. Abdelmalek, Maximum likelihood thresholding based on population of mixture models, *Pattern Recognition* 25 (1992) 1231–1240.  
 [18] C.H. Li, C.K. Lee, Minimum entropy thresholding, *Pattern Recognition* 26 (1993) 617–625.  
 [19] A.D. Brink, Maximum entropy segmentation based on the autocorrelation function of the image histogram, *Comput. Inform. Technol.* 2 (1994) 77–85.  
 [20] L.K. Huang, M.J. Wang, Image thresholding by minimizing the measures of fuzziness, *Pattern Recognition* 28 (1995) 41–51.  
 [21] H.D. Cheng, Y.H. Chen, X.H. Jiang, Thresholding using two-dimensional histogram and fuzzy entropy principle, *IEEE Trans. Image Process.* 9 (2000) 732–735.  
 [22] O.J. Tobias, R. Seara, Image segmentation by histogram thresholding using fuzzy sets, *IEEE Trans. Image Process.* 11 (2002) 1457–1465.  
 [23] S.U. Lee, S.Y. Chung, R.H. Park, A comparative performance study of several global thresholding techniques for segmentation, *Comput. Vision Graph. Image Process.* 52 (1990) 171–190.  
 [24] C.A. Glasbey, An analysis of histogram based thresholding algorithms, *Proc. Graph. Models Image Process.* 55 (1993) 532–537.  
 [25] F. Melgani, G. Moser, S.B. Serpico, Unsupervised change-detection methods for remote sensing images, *Opt. Eng.* 41 (2002) 3288–3297.  
 [26] M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, *J. Electron. Imaging* 13 (2004) 146–168.  
 [27] J.K. Fwu, P.M. Djuric, EM algorithm for image segmentation initialized by a tree structure scheme, *IEEE Trans. Image Process.* 6 (1997) 349–352.  
 [28] M.N. Do, M. Vetterli, Wavelet-based texture retrieval using generalized Gaussian density and Kullback–Leibler distance, *IEEE Trans. Image Process.* 11 (2002) 146–158.  
 [29] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *J. R. Stat. Soc. B* 1 (1977) 1–38.  
 [30] G. McLachlan, T. Krishnan, *The EM Algorithm and Extensions*, Wiley, New York, 1997.  
 [31] M. Abramowitz, I.A. Stegun, *HandBook of Mathematical Tables*, Dover, New York, 1970.  
 [32] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1988.  
 [33] T. Hastie, R. Tibshirani, Discriminant analysis by Gaussian mixtures, *J. R. Stat. Soc. B* 58 (1996) 155–176.

- [34] S. Roberts, D. Husmeier, I. Rezek, W. Penny, Bayesian approaches to Gaussian mixture modeling, *IEEE Trans. Pattern Anal. Mach. Intell.* 20 (1998) 1133–1142.
- [35] G. McLachlan, D. Peel, *Finite Mixture Models*, Wiley, New York, 2000.
- [36] D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, MA, 1989.
- [37] J.H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI, 1975.
- [38] L. Davis, *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York, 1991.
- [39] A. Lopes, R. Touzi, E. Nezry, Adaptive speckle filters and scene heterogeneity, *IEEE Trans. Geosci. Remote Sensing* 28 (1990) 992–1000.
- [40] E.J.M. Rignot, J.J. Van Zyl, Change detection techniques for ERS-1 SAR data, *IEEE Trans. Geosci. Remote Sensing* 31 (1993) 896–906.
- [41] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, Wiley, New York, 2000.