

# Threshold Selection from Image Histograms with Skewed Components Based on Maximum-likelihood Estimation of Skew-normal and Log-concave Distributions

Jing-Hao Xue<sup>a,\*</sup>, D. Michael Titterington<sup>b</sup>

<sup>a</sup>*Department of Statistical Science, University College London, London WC1E 6BT, UK*

<sup>b</sup>*School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8QQ, UK*

---

## Abstract

The purpose of this paper is to propose two new methods for histogram-based image thresholding: one is based on parametric maximum-likelihood estimation of skew-normal distributions, and the other is based on nonparametric maximum-likelihood estimation of log-concave distributions. The main advantages of using these two methods are threefold. First, both methods are natural generalisations of the classical Gaussian-based methods. Secondly, both methods can take into consideration the skewness of the distributions of individual classes. Thirdly, both methods are in line with Otsu's method and the minimum error thresholding (MET) method, based on the comparison of maximum log-likelihoods, for determining an optimal threshold. Compared with Otsu's method and the MET method, the two methods demonstrate comparable, if not better, performance for binarisation of real images and simulated data presented in this paper. Limitations and extensions of the methods are briefly discussed.

*Keywords:* Image thresholding, Log-concave distributions, Minimum error thresholding (MET), Mixture distributions, Maximum-likelihood estimation, Otsu's method, Skew-normal distributions

---

\*Corresponding author. Tel.: +44-20-7679-1863; Fax: +44-20-3108-3105.

*Email addresses:* [jinghao@stats.ucl.ac.uk](mailto:jinghao@stats.ucl.ac.uk) (Jing-Hao Xue),  
[michael.titterington@gla.ac.uk](mailto:michael.titterington@gla.ac.uk) (D. Michael Titterington)

## 1. Introduction

Image thresholding is a simple and widely-used technique for segmentation; it partitions a grey-level image into non-overlapped segments such that each segment belongs to one of some predefined classes, which to some extent can be identified by their grey levels (Sahoo et al., 1988; Pal and Pal, 1993; Sezgin and Sankur, 2004).

Most thresholding approaches are proposed for binarisation of the image into two classes,  $\mathcal{C}_0$  and  $\mathcal{C}_1$ , based on analysis of the grey-level histogram of the image. Typically,  $\mathcal{C}_0$  will be defined by a range of dark grey levels and  $\mathcal{C}_1$  by a complementary range of light grey levels. Also, typically, one of the classes will represent the ‘object’ and the other the ‘background’. Comprehensive overviews and comparative studies of histogram-based image-binarisation approaches can be found in Sahoo et al. (1988), Lee et al. (1990), Glasbey (1993), Trier and Jain (1995) and Sezgin and Sankur (2004), among others.

Two of the most popular approaches are Otsu’s method (Otsu, 1979) and Kittler and Illingworth’s minimum error thresholding (MET) method (Kittler and Illingworth, 1986), both motivated by the results from statistical classification and cluster analysis. More precisely, Otsu’s method was based on Fisher’s linear discriminant analysis, using a within-class variance  $\sigma_W^2$  and a between-class variance  $\sigma_B^2$  (Otsu, 1979). The MET method was based on the Bayes discriminant rule under the assumption that the within-class grey-level distributions  $p(x|y)$  are normal (or Gaussian) distributions, where  $y$  denotes the class indicator of a grey level  $x$  (Kittler and Illingworth, 1986); for example  $y = 0$  for  $x \in \mathcal{C}_0$  and  $y = 1$  for  $x \in \mathcal{C}_1$ . That is, the MET method assumes that the grey levels of the  $N$  pixels in the image compose a sample of  $N$  observations from a two-component Gaussian mixture distribution  $p(x)$ .

Both Otsu’s method and the MET method can be derived from maximisation of the log-likelihoods based on  $p(x|y)$  and  $p(x, y)$ , respectively (Kurita et al., 1992), under certain Gaussian assumptions about  $p(x|y)$ . However, such assumptions may not be perfectly appropriate in some cases of histogram-based image thresholding, in particular when the grey level of the background or the object is far from being normally distributed. Typical of such violation of the Gaussian assumption is a histogram with skewed components; in other words, the grey-level distributions of the background and the object are skewed, which can be very detrimental to the performance of a thresholding method.

38 In order to accommodate data that do not follow Gaussian distributions  
39 for individual classes in some applications, other distributions, such as Pois-  
40 son distributions (Pal and Bhandari, 1993) and certain distributions derived  
41 from Rayleigh (Xue et al., 1999), Nakagami-gamma, Weibull and log-normal  
42 distributions (Moser and Serpico, 2006), have been used to develop variants  
43 of the MET method. These distributions can take the skewness into ac-  
44 count to some extent; however, they do not include Gaussian distributions  
45 as special cases, and thus lose some generality.

46 A generalisation of the Gaussian distribution is the exponential power  
47 distribution, also known as the generalised Gaussian distribution. Recently,  
48 Bazi et al. (2007), Fan et al. (2008) and Fan and Lin (2009) considered using a  
49 finite mixture of generalised Gaussian distributions to fit the image histogram  
50 for thresholding. However, the generalised Gaussian distribution employed in  
51 their work for each component (or class) is a symmetric distribution, although  
52 it covers distributions such as Gaussian, Laplace and uniform distributions.

53 In this context, this paper proposes two new methods for histogram-based  
54 image thresholding: one is based on parametric maximum-likelihood estima-  
55 tion of skew-normal distributions, and the other is based on nonparametric  
56 maximum-likelihood estimation of log-concave distributions.

57 The main advantages of using the skew-normal-based method and the log-  
58 concave-based method are threefold. First, both methods are natural gen-  
59 eralisations of the classical Gaussian-based thresholding methods, because  
60 Gaussian distributions are special cases of skew-normal distributions and  
61 the latter in turn are special cases of log-concave distributions. Secondly,  
62 both methods can take into consideration the skewness of the distributions  
63 of individual classes, since both skew-normal distributions and log-concave  
64 distributions can be skew. Thirdly, we shall show in section 3 that both meth-  
65 ods are in line with Otsu’s method and the MET method, in the sense that  
66 all the methods are based on the comparison of maximum log-likelihoods,  
67 which are associated with different grey-level thresholds  $t$ , for determining  
68 an optimal threshold  $t^*$ .

69 The rest of this paper is organised as follows. Section 2 introduces the  
70 skew-normal distribution and its parametric maximum-likelihood estimation,  
71 and the log-concave distribution and its nonparametric maximum-likelihood  
72 estimation. Section 3 presents four approaches to image thresholding, Otsu’s  
73 method, the MET method and our skew-normal-based and log-concave-based  
74 methods, and their relationships. In section 5, we provide empirical results  
75 obtained from applying these methods to real and simulated data. A discus-

76 sion about extensions and limitations of our methods is given in section 6.

## 77 **2. Skew-normal distributions and log-concave distributions**

### 78 *2.1. Skew-normal distributions*

#### 79 *2.1.1. Definition and some properties*

80 The probability density function  $f_S(X)$  of a random variable  $X$ , which  
81 follows a univariate skew-normal distribution, denoted by  $\text{SN}(\xi, \omega, \alpha)$  here-  
82 after, on  $\mathbb{R}$  with three parameters  $\xi$ ,  $\omega$  and  $\alpha$ , can be written as

$$f_S(X = x) = \frac{2}{\omega} \phi\left(\frac{x - \xi}{\omega}\right) \Phi\left(\alpha \frac{x - \xi}{\omega}\right), \quad (1)$$

83 where  $\phi(\cdot)$  and  $\Phi(\cdot)$  denote the density and distribution functions for the  
84 standard normal  $N(0, 1)$  distribution, respectively. The parameters  $\xi$ ,  $\omega$  and  
85  $\alpha$  are the location, scale and shape parameters, respectively.

86 The properties of univariate and multivariate skew-normal distributions  
87 have been systematically studied by Azzalini (1985), Azzalini and Valle  
88 (1996) and Azzalini and Capitanio (1999), among others. Some properties  
89 desirable for image thresholding are as follows.

90 First, the skew-normal distribution defined in equation (1) includes the  
91 Gaussian distribution  $N(\xi, \omega^2)$  as a special case when the shape parame-  
92 ter  $\alpha$  equals zero. This implies a ‘continuous’ transition from normality to  
93 non-normality (Azzalini, 1985). Secondly, when  $\alpha = 0$ , the  $N(\xi, \omega^2)$  distri-  
94 bution has no skewness; when  $|\alpha|$  increases, the skewness (in absolute value)  
95 increases. In addition, the sign of  $\alpha$  determines the direction of the skew-  
96 ness of  $f_S(X = x)$ . This implies that the shape parameter  $\alpha$  regulates the  
97 skewness of the skew-normal distribution. Thirdly,  $f_S(X = x)$  is strongly  
98 uni-modal, i.e.,  $\log f_S(X = x)$  is a concave function of  $x$ . In other words,  
99 the skew-normal distribution is a log-concave distribution and thus also en-  
100 joys the properties of the latter, which will be discussed in more detail in  
101 section 2.2.

#### 102 *2.1.2. Parametric maximum-likelihood estimation*

103 Based on the parametric expression of the skew-normal distribution as in  
104 equation (1), and given a sample of  $N_y$  identically, independently distributed  
105 (i.i.d.) observations  $\{x_i\}_{i=1}^{N_y}$  from class  $\mathcal{C}_y$ , the log-likelihood function of the

106 parameters  $\theta_S = (\xi, \omega, \alpha)$  can be written as

$$\ell(\theta_S) = -N_y \log \omega - \frac{1}{2} \sum_{i=1}^{N_y} x_i^2 + \sum_{i=1}^{N_y} \log \left\{ 2\Phi \left( \alpha \frac{x_i - \xi}{\omega} \right) \right\} . \quad (2)$$

107 Due to the last term on the right-hand side of equation (2), there is no  
 108 closed-form solution to the maximum-likelihood estimation of  $\theta_S$  (except for  
 109 the Gaussian distribution, i.e.  $\alpha = 0$ ), so numerical computation is usually  
 110 adopted.

111 The parameterisation of  $f_S(X = x)$  given by  $(\xi, \omega, \alpha)$  as in equation (1) is  
 112 easy to understand and convenient for interpretation of the parameters. How-  
 113 ever, with such parameterisation, the Fisher information matrix is singular  
 114 at  $\alpha = 0$  and the shape of the log-likelihood can be far from being quadratic.  
 115 Therefore, in order to overcome these problems, Azzalini (1985) proposes  
 116 a centred parameterisation, which is based on the standardised  $\text{SN}(0, 1, \alpha)$   
 117 distribution instead of the non-standardised  $\text{SN}(0, 1, \alpha)$  itself. These two  
 118 schemes of parameterisation can be readily converted to each other. For sim-  
 119 plicity, we shall omit the centred parameterisation here, and refer to Azzalini  
 120 and Capitanio (1999) for more discussion of parameterisation and other issues  
 121 related to the parametric maximum-likelihood estimation of the skew-normal  
 122 distribution.

## 123 2.2. Log-concave distributions

### 124 2.2.1. Definition and some properties

125 The probability density function  $f_L(X)$  of a random variable  $X$ , which  
 126 follows a univariate log-concave distribution on  $\mathbb{R}$ , can be written as

$$f_L(X = x) = \exp\{\phi_c(x)\} , \quad (3)$$

127 where  $\phi_c(x) : \mathbb{R} \rightarrow [-\infty, \infty)$  is a concave function of  $x$ .

128 Log-concave distributions include many commonly-used distributions, such  
 129 as Gaussian, Laplace, logistic distributions and some gamma, beta and Weibull  
 130 distributions (Walther, 2009; Cule et al., 2010). Log-concave distributions  
 131 can be skew, and skew normal distributions are special cases of log-concave  
 132 distributions (Azzalini, 1985; Gupta and Balakrishnan, 2010).

133 As a result of the nonparametric way they are defined, log-concave distri-  
 134 butions are quite flexible. In addition, log-concave distributions enjoy many  
 135 desirable properties, such as that they are invariant to convolution, product  
 136 and marginalisation and that their densities are uni-modal.

137 Therefore, nonparametric maximum-likelihood estimation for log-concave  
 138 distributions, or more generally shape-constrained nonparametric density es-  
 139 timation, has attracted considerable research interest recently on theoretical,  
 140 methodological and computational aspects; see a recent review by Walther  
 141 (2009), for example.

142 *2.2.2. Nonparametric maximum-likelihood estimation*

143 Based on the nonparametric expression in equation (3), and given i.i.d.  
 144 observations  $\{x_i\}_{i=1}^{N_y}$  from the class  $\mathcal{C}_y$ , the log-likelihood function of  $\phi_c(\cdot)$  is

$$\ell(\phi_c) = \sum_{i=1}^{N_y} \phi_c(x_{(i)}) , \quad (4)$$

145 where  $x_{(i)}$  are the ordered observations,  $x_{(1)} < \dots < x_{(N_y)}$ , for optimisation  
 146 convenience.

147 It is not a trivial matter to maximise the log-likelihood function in equa-  
 148 tion (4) over all concave functions  $\phi_c(\cdot)$ , as the function  $\phi_c(\cdot)$  is subject to  
 149 the constraints that it is a concave function and its exponential is a density  
 150 function (i.e.  $\int_{\mathbb{R}} \exp\{\phi_c(x)\} = 1$ ). This leads to a constrained optimisation  
 151 problem, namely to maximise numerically the following objective function:

$$\frac{1}{N_y} \sum_{i=1}^{N_y} \phi_c(x_{(i)}) - \int_{[x_{(1)}, x_{(N_y)}]} \exp\{\phi_c(x)\} dx , \quad (5)$$

152 where  $\int_{[x_{(1)}, x_{(N_y)}]}$  replaces  $\int_{\mathbb{R}}$  for computation of nonparametric maximum-  
 153 likelihood estimation.

154 This shape-constrained optimisation problem can be tackled by the use  
 155 of specific optimisation algorithms, such as an iterative convex minorant  
 156 algorithm (ICMA) by Walther (2002) and Rufibach (2007) among others, an  
 157 active set algorithm (ASA) by Dümbgen et al. (2007) and a version of Shor's  
 158  $r$ -algorithm by Cule et al. (2010).

159 The nonparametric maximum-likelihood estimation of a log-concave den-  
 160 sity  $f_L(x)$  results in a unique estimator  $\hat{\phi}_c(x)$  of the concave function  $\phi_c(x)$ ,  
 161 where  $\hat{\phi}_c(x)$  is piecewise linear with knots at a subset of the i.i.d. observa-  
 162 tions  $\{x_i\}_{i=1}^{N_y}$ , and vanishes outside the empirical support of  $[x_{(1)}, x_{(N_y)}]$  for  
 163  $\mathcal{C}_y$  (Dümbgen and Rufibach, 2010).

164 **3. Methods for selecting an optimal threshold**

165 In this section, we shall first provide the original rules underlying Otsu's  
 166 method and the MET method for selecting an optimal threshold  $t^*$ , then  
 167 present these two rules again but from the perspective of comparison of max-  
 168 imal likelihoods, and finally propose the rule underlying our skew-normal-  
 169 based and log-concave-based methods for image binarisation.

170 The procedure of image binarisation can be described as follows. For  
 171 an image  $\mathcal{X}$  of  $N$  pixels, each pixel is represented by its grey level  $x_i, i =$   
 172  $1, \dots, N$ . A grey-level threshold  $t$  partitions the image into two classes  $\mathcal{C}_0(t)$   
 173 and  $\mathcal{C}_1(t)$ , where  $\mathcal{C}_0(t) = \{i : 0 \leq x_i < t, 1 \leq i \leq N\}$  and  $\mathcal{C}_1(t) = \{i : t \leq$   
 174  $x_i \leq T, 1 \leq i \leq N\}$ , where  $T$  is the largest possible grey level of  $\mathcal{X}$  (e.g.  
 175  $T = 255$  for 8-bit grey-level images). That is,  $\mathcal{C}_0(t)$  includes all the pixels  
 176 with grey levels  $x$  less than  $t$ , and  $\mathcal{C}_1(t)$  consists of the remaining pixels.

177 *3.1. Otsu's method and the MET method*

178 With thresholding in mind, Otsu (1979) introduces three measures:  $\sigma_B^2(t)/\sigma_W^2(t)$ ,  
 179  $\sigma_T^2(t)/\sigma_W^2(t)$  and  $\sigma_B^2(t)/\sigma_T^2(t)$ , of which  $\sigma_W^2(t)$  is the (sample) within-class  
 180 variance,  $\sigma_B^2(t)$  is the (sample) between-class variance and  $\sigma_T^2$  denotes the  
 181 total variance such that  $\sigma_T^2 = \sigma_W^2(t) + \sigma_B^2(t)$ , for a candidate  $t$ . The to-  
 182 tal variance  $\sigma_T^2$  does not change with the threshold  $t$  whereas the other two  
 183 variances do. Otsu (1979) suggests selecting the optimal threshold  $t^*$  as

$$t_O^* = \operatorname{argmax}_t \sigma_B^2(t) = \operatorname{argmax}_t [\pi_0(t)\pi_1(t) \{\mu_1(t) - \mu_0(t)\}^2] , \quad (6)$$

184 where  $\pi_0(t)$  is the proportion and  $\mu_0(t)$  is the sample mean of class  $\mathcal{C}_0(t)$ ;  
 185  $\pi_1(t)$  and  $\mu_1(t)$  are defined similarly for  $\mathcal{C}_1(t)$ . As  $\sigma_T^2 = \sigma_W^2(t) + \sigma_B^2(t)$  is  
 186 invariant to  $t$ , the criterion in equation (6) is equivalent to the selection of  $t^*$   
 187 as

$$t_O^* = \operatorname{argmin}_t \sigma_W^2(t) = \operatorname{argmin}_t \{ \pi_0(t)\sigma_0^2(t) + \pi_1(t)\sigma_1^2(t) \} , \quad (7)$$

188 where  $\sigma_0(t)$  and  $\sigma_1(t)$  are the sample standard deviations within  $\mathcal{C}_0(t)$  and  
 189  $\mathcal{C}_1(t)$ , respectively.

190 Kittler and Illingworth's MET method (Kittler and Illingworth, 1986)  
 191 selects  $t^*$  as

$$t_M^* = \operatorname{argmin}_t \left\{ \pi_0(t) \log \frac{\sigma_0(t)}{\pi_0(t)} + \pi_1(t) \log \frac{\sigma_1(t)}{\pi_1(t)} \right\} , \quad (8)$$

192 where in practice usually only a  $t$  with positive  $\pi_0(t)$ ,  $\pi_1(t)$ ,  $\sigma_0(t)$  and  $\sigma_1(t)$   
 193 is a candidate threshold.

194 Otsu’s method and the MET method are two of the most widely-used  
 195 approaches to image thresholding. Research efforts have been made to in-  
 196 terpret and unify them from various perspectives (Kurita et al., 1992; Yan,  
 197 1996; Xue and Titterington, 2010b).

198 Kurita et al. (1992) show that the Gaussian-based Otsu method is equiv-  
 199 alent to the search for the threshold  $t$  that provides the largest maximum log-  
 200 likelihood based on  $p(x|y; t)$ , under the assumption that  $x|y \sim N(\mu_y(t), \sigma_y^2(t))$   
 201 represents the within-class Gaussian distributions with the class indicator  $y$ ,  
 202 and  $\sigma_0^2(t) = \sigma_1^2(t)$  for any candidate  $t$ . That is, the Gaussian-based Otsu  
 203 method selects an optimal threshold  $t_O^*$  that satisfies

$$t_O^* = \operatorname{argmax}_t \left[ \max_{\theta_O(t)} \left\{ \sum_{i=1}^N \log p(x_i|y_i; t) \right\} \right], \quad (9)$$

204 where population parameters  $\theta_O(t) = (\mu_0(t), \sigma_0^2(t), \mu_1(t), \sigma_1^2(t))$  are estimated  
 205 through maximisation of the log-likelihood (i.e. by their sample estimators).  
 206 By abuse of notation, we use the same symbols for population parameters  
 207 and their sample estimators, when there is no ambiguity in the context.

208 Kurita et al. (1992) also show that Kittler and Illingworth’s MET method  
 209 is equivalent to the search for the threshold  $t$  that provides the largest max-  
 210 imum log-likelihood based on the joint distribution  $p(x, y; t)$ , under the as-  
 211 sumptions that  $x|y \sim N(\mu_y(t), \sigma_y^2(t))$  and  $\sigma_0^2(t)$  is not necessarily to equal  
 212  $\sigma_1^2(t)$  for any candidate  $t$ . In addition,  $p(x, y; t)$  is factorised as  $p(y; t)p(x|y; t)$ .  
 213 That is, the threshold selection rule is

$$t_M^* = \operatorname{argmax}_t \left[ \max_{\theta_M(t)} \left\{ \sum_{i=1}^N \log \{p(y_i; t)p(x_i|y_i; t)\} \right\} \right], \quad (10)$$

214 where  $\theta_M(t) = (\pi_0(t), \mu_0(t), \sigma_0^2(t), \mu_1(t), \sigma_1^2(t))$ , in which  $\pi_0(t) = p(y(t) = 0)$   
 215 is the class prior for a candidate threshold  $t$ .

### 216 3.2. Our skew-normal-based and log-concave-based methods

217 As with the maximum-likelihood estimation of Gaussian distributions in  
 218 Otsu’s method and the MET method (Kurita et al., 1992) for image thresh-  
 219 olding, the maximum-likelihood estimation of skew-normal distributions and

220 log-concave distributions can also provide a maximal likelihood for each can-  
 221 didate threshold  $t$ . These maximal likelihoods can then be compared so as  
 222 to select an optimal threshold  $t^*$ , which leads to our skew-normal-based and  
 223 log-concave-based methods for image thresholding.

224 This characteristic, together with skewness, uni-modality, flexibility and  
 225 generality (in particular in including Gaussian distributions as special cases)  
 226 of skew-normal distributions and log-concave distributions, makes the para-  
 227 metric or nonparametric maximum-likelihood estimation of these distribu-  
 228 tions appropriate for histogram-based image thresholding.

229 Our skew-normal-based and log-concave-based methods use the same rule  
 230 as that in equation (10) for selecting an optimal threshold  $t^*$ , except that we  
 231 assume a skew-normal distribution  $f_S(x|y; t)$  or a log-concave distribution  
 232  $f_L(x|y; t)$  instead of a Gaussian distribution for each class  $\mathcal{C}_y(t)$  and any  
 233 candidate  $t$ .

234 Therefore, our skew-normal-based rule for the selection of  $t^*$  is

$$t_S^* = \operatorname{argmax}_t \left[ \sum_{i=1}^N \log\{\pi_{y_i}(t) \hat{f}_S(x_i|y_i; t)\} \right], \quad (11)$$

235 where, for a candidate  $t$ ,  $\hat{f}_S(x|y; t)$  is the parametric maximum-likelihood  
 236 estimator of the skew-normal density of class  $\mathcal{C}_y(t)$  determined by  $t$ , and  
 237  $\pi_y(t)$  is the proportion of pixels in  $\mathcal{C}_y(t)$ . Similarly, our log-concave-based  
 238 rule for the selection of  $t^*$  is

$$t_L^* = \operatorname{argmax}_t \left[ \sum_{i=1}^N \log\{\pi_{y_i}(t) \hat{f}_L(x_i|y_i; t)\} \right], \quad (12)$$

239 where  $\hat{f}_L(x|y; t)$  is the nonparametric maximum-likelihood estimator of the  
 240 log-concave density for class  $\mathcal{C}_y(t)$  determined by  $t$ .

#### 241 4. Implementation of the methods

242 The image-thresholding methods compared in this work were implemented  
 243 in R, a free software environment for statistical computing and graphics  
 244 ([www.r-project.org](http://www.r-project.org)).

245 Our implementation, of Otsu's method, the MET method, the skew-  
 246 normal-based method and the log-concave-method, is based on equations

247 (7), (8), (11) and (12), respectively. For purposes of consistency, comparison  
 248 and illustration, we rewrite the equations briefly as

$$t_O^* = \underset{t}{\operatorname{argmin}} J_O(t) , \quad (13)$$

$$t_M^* = \underset{t}{\operatorname{argmin}} J_M(t) , \quad (14)$$

$$t_S^* = \underset{t}{\operatorname{argmin}} J_S(t) , \quad (15)$$

$$t_L^* = \underset{t}{\operatorname{argmin}} J_L(t) , \quad (16)$$

249 where  $J_O(t)$  and  $J_M(t)$  are the terms to be minimised on the right-hand sides  
 250 of equations (7) and (8), and  $J_S(t)$  and  $J_L(t)$  are the additive inverses of the  
 251 terms to be maximised on the right-hand sides of equations (11) and (12).

252 For Otsu's method and the MET method, the maximum-likelihood esti-  
 253 mators, namely the proportion, sample mean and sample variance for each  
 254 class, have closed-form solutions and thus are readily implemented. In con-  
 255 trast, for our skew-normal-based and log-concave-based methods, there are  
 256 no closed-form solutions for the estimation, and thus additional consider-  
 257 ation, as discussed below, is needed in order to accelerate the numerical  
 258 computation.

#### 259 4.1. Our skew-normal-based method

260 The implementation of parametric maximum-likelihood estimation of uni-  
 261 variate and multivariate skew-normal distributions can be found in an R  
 262 package `sn`.

263 In the univariate case, with centred parameterisation, a gradient-based  
 264 optimisation algorithm is used for maximum-likelihood estimation; alterna-  
 265 tively, with direct parameters  $(\xi, \omega, \alpha)$ , an expectation-maximisation (EM)-  
 266 type algorithm is used. Both algorithms are very slow when we treat each  
 267 pixel separately, as there are  $N_y$  pixels for class  $\mathcal{C}_y$  where  $N_y$  is in general a  
 268 large number; for example, for an image of  $256 \times 256$  pixels of two classes  $\mathcal{C}_0$   
 269 and  $\mathcal{C}_1$ ,  $N = N_0 + N_1 = 65536$ .

270 Fortunately, based on their grey levels,  $N$  pixels can be treated as grouped  
 271 data, such that there are at most  $T+1$  groups and each group has  $h(x)$  pixels  
 272 (i.e., the frequency of grey level  $x$  is  $h(x)$  as displayed in the histogram of  
 273 the image). Hence, for class  $\mathcal{C}_0(t)$ , the log-likelihood function in equation (2)

274 can be rewritten as

$$\ell(\theta_S; t) = -N_0 \log \omega - \frac{1}{2} \sum_{x=0}^{t-1} \{h(x)x^2\} + \sum_{x=0}^{t-1} \left[ h(x) \log \left\{ 2\Phi \left( \alpha \frac{x - \xi}{\omega} \right) \right\} \right] , \quad (17)$$

275 where  $N_0 = \sum_{x=0}^{t-1} h(x)$ . The formula for class  $\mathcal{C}_1$  can be derived simi-  
 276 larly. This substantially improves the computational speed of the maximum-  
 277 likelihood estimation of the skew-normal distribution. Meanwhile, such com-  
 278 putation does not slow down when the image size becomes larger.

279 Therefore, in this study, we adopt a routine called *msn.mle* in the pack-  
 280 age **sn** for the estimation. This routine can deal with univariate grouped  
 281 data; it uses the method of moments to initialise parameters and a general  
 282 optimisation procedure, *nlnmb*, to maximise the log-likelihood; it is fast and  
 283 robust in our experiments for image thresholding, although it may not con-  
 284 verge when  $t$  is at the ends of the grey-level range (i.e., when a class has few  
 285 grey levels).

#### 286 4.2. Our log-concave-based method

287 Two implementations of nonparametric maximum-likelihood estimation  
 288 of univariate log-concave densities can be found in an R package **logcondens**  
 289 (Dümbgen and Rufibach, 2010), by using an ICMA and an ASA, respectively.  
 290 In this study, we use the implementation based on ASA, the most efficient  
 291 method to date (Walther, 2009). In the multivariate case, an implementation  
 292 can be found in an R package **LogConcDEAD** (Cule et al., 2009), which also  
 293 provides the univariate estimation but, unlike those of **logcondens**, by using  
 294 a version of Shor's  $r$ -algorithm for optimisation.

295 As with the skew-normal-based method, using grey levels we can treat  
 296 the pixels as grouped data such that the computational load is substantially  
 297 reduced. More specifically, for class  $\mathcal{C}_0(t)$ , the log-likelihood function of  $\phi_c(\cdot)$   
 298 in equation (4) can be rewritten as

$$\ell(\phi_c; t) = \sum_{x=0}^{t-1} \{h(x)\phi_c(x)\} , \quad (18)$$

299 and the objective function in equation (5) can be rewritten as

$$\frac{1}{N_0} \sum_{x=0}^{t-1} \{h(x)\phi_c(x)\} - \int_0^{t-1} \exp\{\phi_c(x)\} dx . \quad (19)$$

300 Similar formulae can be derived for class  $\mathcal{C}_1(t)$ . The computational speed is  
 301 invariant to the image size.

302 Some discussion about the limitations of the nonparametric maximum-  
 303 likelihood estimation of log-concave densities can be found in Walther (2009),  
 304 Dümbgen and Rufibach (2010) and in particular Cule et al. (2010). The  
 305 computation for nonparametric estimation is in general much slower than  
 306 that for its parametric counterpart; computational speed is also an issue for  
 307 our log-concave-based thresholding method.

308 In our experiments, with no additional effort being made to optimise the  
 309 computer program, the time taken by Otsu’s method, the MET method, our  
 310 skew-normal-based and log-concave-based methods is approximately in the  
 311 ratio 1:1:8:22 for some real images of  $256 \times 256$  pixels and in the ratio 1:1:2:6  
 312 for some real images of  $512 \times 512$  pixels.

## 313 5. Results

314 For illustrative purposes, in this section we apply the four binarisation  
 315 methods to four real images and twenty-five simulated data sets.

### 316 5.1. Real images

317 The four real images that we chose are ‘NDT-image5’, ‘NDT-image19’,  
 318 ‘NDT-image20’ and ‘Coins’. ‘NDT-image5’, ‘NDT-image19’ and ‘NDT-image20’  
 319 are three nondestructive testing (NDT) images adopted in a recent compre-  
 320 hensive survey (Sezgin and Sankur, 2004) of image-binarisation methods.  
 321 ‘Coins’ was used by the MATLAB Image Processing Toolbox (The Math-  
 322 Works, Inc.) for demonstration of Otsu’s method and by some studies of  
 323 image thresholding including one of our recent pieces of work (Xue and Tit-  
 324 ington, 2010a).

	$t_O^*$	$t_M^*$	$t_S^*$	$t_L^*$
NDT-image5	150	89	91	224
NDT-image19	167	194	190	204
NDT-image20	80	39	40	47
Coins	127	95	79	79

Table 1: Thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$  selected by the use of Otsu’s method, the MET method and our skew-normal-based and log-concave-based methods, respectively, for four real images.

325 The values of  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$  selected by the four methods are listed  
 326 in Table 1. In Figures 1–4, we show for each of the four images its original  
 327 version, its histogram with thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$  superimposed and  
 328 indicated by dashed lines, the curves for  $J_O(t)$ ,  $J_M(t)$ ,  $J_S(t)$  and  $J_L(t)$  (all of  
 329 which have been rescaled to the range  $[0, 1]$ ), and its binarised images based  
 330 on  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$ , respectively.

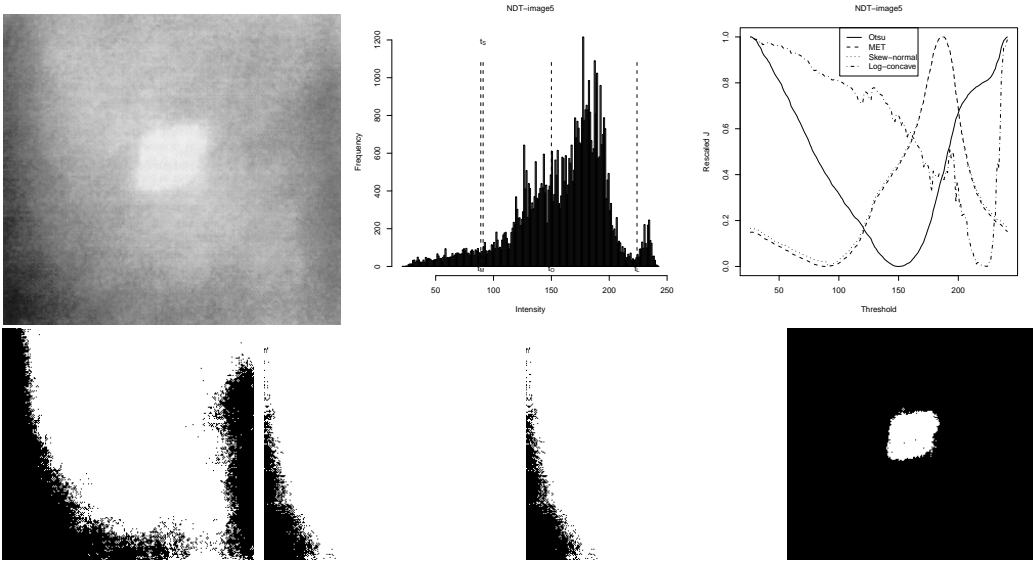


Figure 1: For ‘NDT-image5’. Upper row, from left to right: original image; histogram with thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$  superimposed and indicated by dashed lines;  $J_O(t)$ ,  $J_M(t)$ ,  $J_S(t)$  and  $J_L(t)$ , all of which have been rescaled to the range  $[0, 1]$ . Lower row, from left to right: binarised images based on  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$ , respectively. Binarised images are scaled down.

331 In the centre of ‘NDT-image5’, an infrared thermal image for defect de-  
 332 tection as shown in Figure 1, the bright hole (as  $\mathcal{C}_1$ ) is of much smaller size  
 333 and its grey level is of much smaller variance, compared with the inhomoge-  
 334 neous background (as  $\mathcal{C}_0$ ). In this case, only the log-concave-based method  
 335 achieves a reasonable threshold (i.e.  $t_L^*$ ), which is located in the histogram  
 336 valley between  $\mathcal{C}_0$  and  $\mathcal{C}_1$ , while other three methods fail to extract the hole.

337 ‘NDT-image19’ is a light microscope image of a material’s microstructure.  
 338 As shown by its histogram in Figure 2, the grey-to-dark material ( $\mathcal{C}_0$ ) and the  
 339 bright background ( $\mathcal{C}_1$ ) are of similar sizes but with different variance in grey

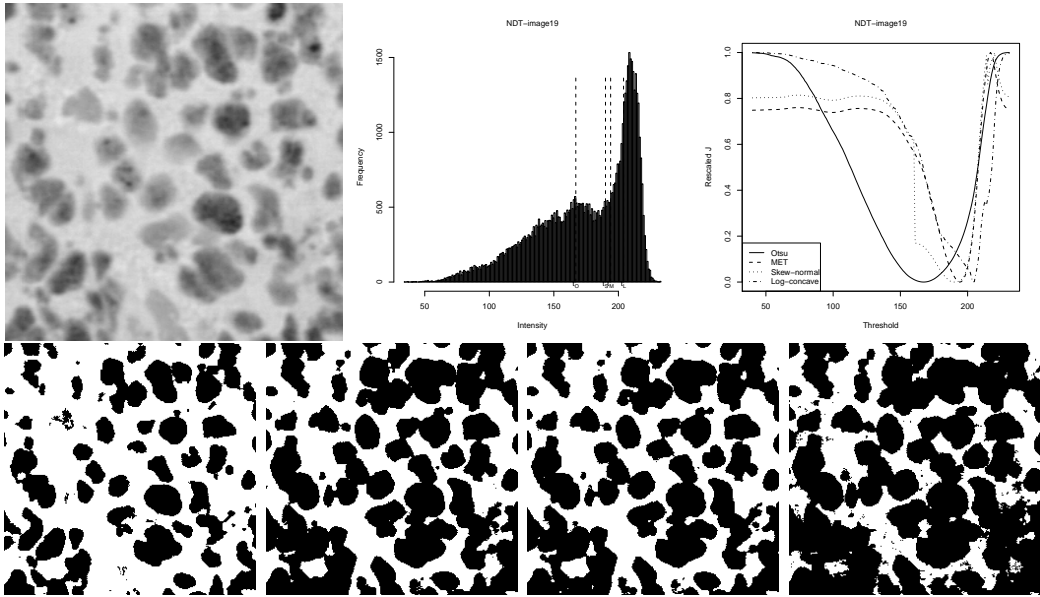


Figure 2: For ‘NDT-image19’. Caption is as for Figure 1.

340 level. Although the histogram for  $\mathcal{C}_0$  can be regarded as being skew, which  
 341 leads to different curves for  $J_M(t)$  and  $J_S(t)$ ,  $t_M^*$  and  $t_S^*$  are similarly located  
 342 between the histogram modes for  $\mathcal{C}_0$  and  $\mathcal{C}_1$ , both producing nearly optimal  
 343 binarisation results. Otsu’s method and the log-concave-based method do  
 344 not perform well in this case:  $t_O^*$  is too small while  $t_L^*$  is too large.

345 ‘NDT-image20’ is another light microscope image intended to inspect a  
 346 material’s microstructure; ‘Coins’ is an ordinary picture. As shown in Fig-  
 347 ures 3 and 4, the histograms of these two images are of different types: for  
 348 ‘NDT-image20’, the class of black pores ( $\mathcal{C}_0$ ) has a smaller size and more  
 349 compact grey levels than the class of grey particles ( $\mathcal{C}_1$ ); for ‘Coins’, the  
 350 dark background ( $\mathcal{C}_0$ ) contains many more pixels, and has much more com-  
 351 pact, skewed grey levels, than the grey coins ( $\mathcal{C}_1$ ). In both cases, the MET,  
 352 the skew-normal-based and the log-concave-based methods perform similarly  
 353 well, much better than does Otsu’s method. All thresholds,  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$ ,  
 354 are located between the histogram modes for  $\mathcal{C}_0$  and  $\mathcal{C}_1$ , but  $t_O^*$  is relatively  
 355 large.

356 We note that, in our experiments, the curve for  $J_S(t)$  is often close to that  
 357 for  $J_M(t)$ , as shown in Figures 1–4. This implies that, for each of the four

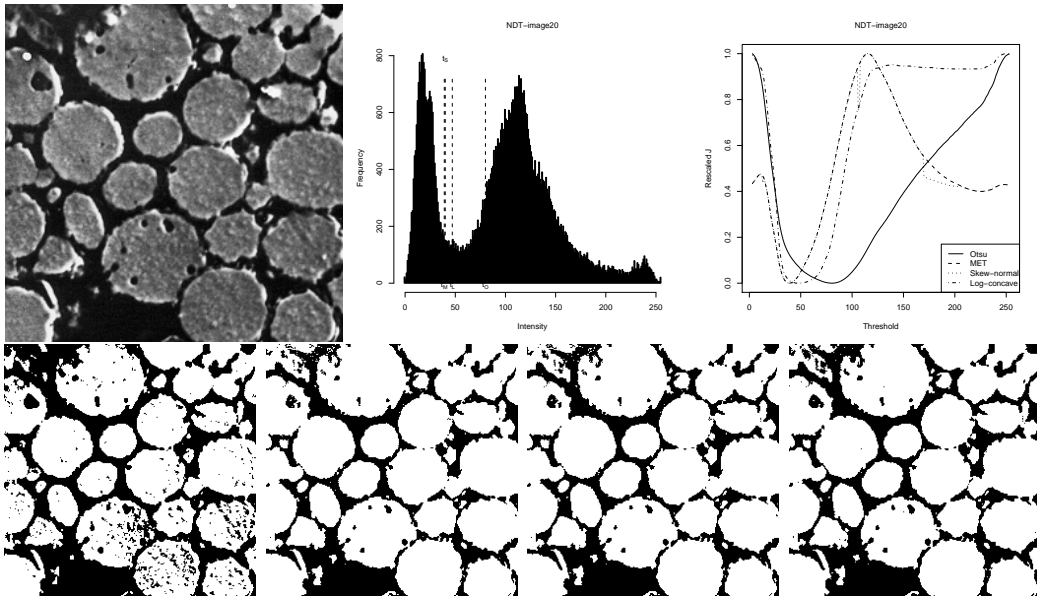


Figure 3: For 'NDT-image20'. Caption is as for Figure 1.

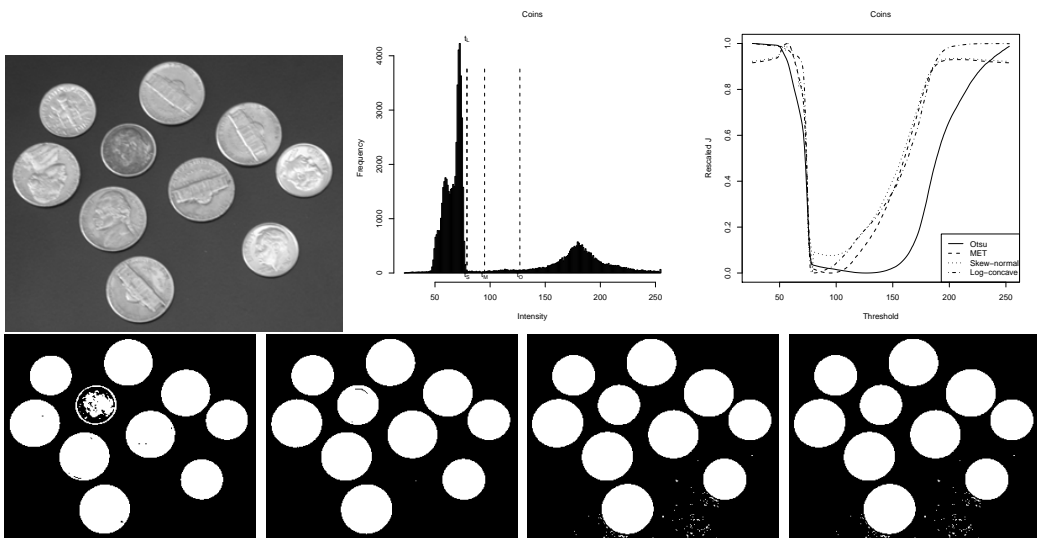


Figure 4: For 'Coins'. Caption is as for Figure 1.

358 images, the value of the shape parameter  $\alpha$  of the skew-normal distribution  
 359 obtained from the maximum-likelihood estimation is often approximately  
 360 zero (which corresponds to a nearly Gaussian distribution), except when the  
 361 histogram for a class is remarkably skew as shown in Figures 2 and 4.

## 362 5.2. Simulated histograms

363 Although some patterns may be observed from empirical studies by using  
 364 real images such as those presented in section 5.1, it is helpful to perform ad-  
 365 ditional studies on simulated data sets, in order to reach some more general-  
 366 isable, reliable and interpretable patterns and to provide a good complement  
 367 to the empirical studies.

368 Therefore, in this section, we present the results obtained from apply-  
 369 ing the four binarisation methods to 25 sets of simulated data from two-  
 370 component skew-normal mixtures. In each set, there are  $N = 256 \times 256$  data  
 371 points, and the data correspond to grey levels of all pixels in a virtual image  
 372  $\mathcal{X}$  of two classes,  $\mathcal{C}_0$  and  $\mathcal{C}_1$ , with class proportions  $\pi_0$  and  $\pi_1$ . For  $\mathcal{C}_0$ , the  
 373 grey level follows a skew-normal  $\text{SN}(\xi_0, \omega_0, \alpha_0)$  distribution; for  $\mathcal{C}_1$ , the grey  
 374 level follows  $\text{SN}(\xi_1, \omega_1, \alpha_1)$ . All simulated data are rounded, left-truncated  
 375 and right-censored into the range  $[0, T]$ , where  $T = 255$ , as grey levels are  
 376 non-negative integers.

377 The reasons why we build 25 sets of simulated data are as follows.

378 First, we consider five types of skewness for a two-component mixture:  
 379 one type is for the case that neither component is skew (i.e. with  $\alpha_1 =$   
 380  $\alpha_2 = 0$ ) corresponding to two Gaussian components, denoted by subscripts  
 381  $G$  hereafter; the other four types are for cases in which both components are  
 382 skew, either left-skewed (i.e. with negative  $\alpha_y$ ) or right-skewed (i.e. with  
 383 positive  $\alpha_y$ ), denoted hereafter by subscripts  $LL$ ,  $RR$ ,  $LR$  and  $RL$  where  
 384  $R$  and  $L$  represents right-skewed and left-skewed, respectively. Here we do  
 385 not consider cases in which only a single component is skew, which are less  
 386 complicated than cases in which both components are skew.

387 Secondly, we consider the different relationships, namely larger than,  
 388 equal to and smaller than, between proportions  $\pi_0$  and  $\pi_1$  and between stan-  
 389 dard deviations  $\sigma_0$  and  $\sigma_1$  for  $\mathcal{C}_0$  and  $\mathcal{C}_1$ . There are nine combinations of the  
 390 pairwise relationships, but only five of them are unique since some pairs are  
 391 mirrors of each other (Xue and Titterington, 2010a). The data sets corre-  
 392 sponding to these five unique combinations are denoted by  $\mathcal{X}_1$ ,  $\mathcal{X}_2$ ,  $\mathcal{X}_3$ ,  $\mathcal{X}_4$   
 393 and  $\mathcal{X}_5$ , respectively; their characteristics can be found in Table 2.

394 Therefore, in this paper the 25 data sets are denoted by  $\mathcal{X}_{1G}, \dots, \mathcal{X}_{5RL}$ ,  
 395 respectively.

396 Furthermore, as with Glasbey (1993), Xue and Titterington (2010b) and  
 397 Xue and Titterington (2010a), we set  $\mu_0 = 100$  and  $\mu_1 = 140$  for  $\mathcal{C}_0$  and  
 398  $\mathcal{C}_1$ . We set  $\alpha_y = \pm 4$  for the skew-normal distributions. Then, based on the  
 399 properties of the skew-normal distribution that

$$\mu = \xi + \omega\delta\sqrt{2/\pi}, \quad (20)$$

$$\sigma = \omega\sqrt{1 - 2\delta^2/\pi}, \quad (21)$$

400 where  $\pi$  is Archimedes' constant and  $\delta = \alpha/\sqrt{1 + \alpha^2}$ , we can obtain  $\omega_y$  and  
 401  $\xi_y$  for class  $\mathcal{C}_y$ , as listed in Tables 3–7 rounded to the nearest integers.

	$\pi_0$	$\pi_1$	$\sigma_0$	$\sigma_1$
$\mathcal{X}_1$	0.5	0.5	10	10
$\mathcal{X}_2$	0.5	0.5	5	15
$\mathcal{X}_3$	0.8	0.2	10	10
$\mathcal{X}_4$	0.8	0.2	5	15
$\mathcal{X}_5$	0.8	0.2	15	5

Table 2: Five combinations of relationships between proportions  $\pi_0$  and  $\pi_1$  and between standard deviations  $\sigma_0$  and  $\sigma_1$  used for data simulation in this paper.

	$\alpha_0$	$\alpha_1$	$\omega_0$	$\omega_1$	$\xi_0$	$\xi_1$	$t_O^*$	$t_M^*$	$t_S^*$	$t_L^*$
$\mathcal{X}_{1G}$	0	0	10	10	100	140	120	120	120	117
$\mathcal{X}_{1LL}$	-4	-4	16	16	112	152	121	125	123	116
$\mathcal{X}_{1RR}$	4	4	16	16	88	128	120	116	122	124
$\mathcal{X}_{1LR}$	-4	4	16	16	112	128	120	120	120	123
$\mathcal{X}_{1RL}$	4	-4	16	16	88	152	121	121	121	123

Table 3: The shape parameters ( $\alpha_0$  and  $\alpha_1$ ), scale parameters ( $\omega_0$  and  $\omega_1$ ) and location parameters ( $\xi_0$  and  $\xi_1$ ) of two-component skew-normal mixtures for five simulated data sets ( $\mathcal{X}_{1G}$ ,  $\mathcal{X}_{1LL}$ ,  $\mathcal{X}_{1RR}$ ,  $\mathcal{X}_{1LR}$  and  $\mathcal{X}_{1RL}$ ), with corresponding thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$ . Note that  $\mathcal{X}_{1G}$  corresponds to a two-component Gaussian mixture.

402 The histograms constructed from the first set of five data sets,  $\mathcal{X}_{1G}, \dots, \mathcal{X}_{1RL}$ ,  
 403 are plotted in Figure 5, with corresponding thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$  su-  
 404 perimposed. From Figure 5, we can observe that, in this case of two classes  
 405 with equal sizes, equal variances but different types of skewness, all four

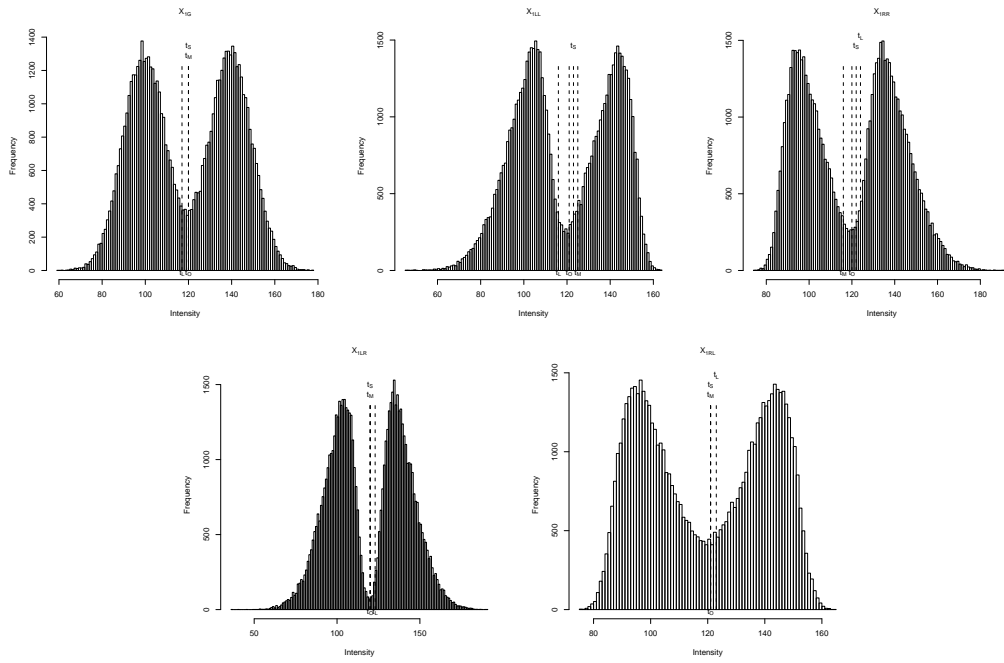


Figure 5: Histograms for  $\mathcal{X}_{IG}$ ,  $\mathcal{X}_{ILL}$ ,  $\mathcal{X}_{IRR}$ ,  $\mathcal{X}_{ILR}$  and  $\mathcal{X}_{1RL}$ , with thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$  indicated by dashed lines.

406 methods provide reasonable thresholds, well located in the valley between  
 407 the two classes, although, when the directions of the skewness of the two  
 408 classes are the same, some thresholds are not perfectly located at the lowest  
 409 point of the valley.

	$\alpha_0$	$\alpha_1$	$\omega_0$	$\omega_1$	$\xi_0$	$\xi_1$	$t_O^*$	$t_M^*$	$t_S^*$	$t_L^*$
$\mathcal{X}_{2G}$	0	0	5	15	100	140	123	113	113	110
$\mathcal{X}_{2LL}$	-4	-4	8	24	106	159	124	115	115	109
$\mathcal{X}_{2RR}$	4	4	8	24	94	121	122	110	112	111
$\mathcal{X}_{2LR}$	-4	4	8	24	106	121	122	112	108	109
$\mathcal{X}_{2RL}$	4	-4	8	24	94	159	124	114	114	112

Table 4: The shape, scale and location parameters for  $\mathcal{X}_{2G}$ ,  $\mathcal{X}_{2LL}$ ,  $\mathcal{X}_{2RR}$ ,  $\mathcal{X}_{2LR}$  and  $\mathcal{X}_{2RL}$ , with corresponding thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$ .

410 In the second set of five data sets,  $\mathcal{X}_{2G}, \dots, \mathcal{X}_{2RL}$ , the two classes are of  
 411 equal sizes but with remarkably different variances. As shown in Figure 6,  
 412 in this case,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$  are in general well located in the valley, while  $t_O^*$   
 413 moves towards the class with larger variance. Roughly speaking, the case of  
 414 ‘NDT-image19’, as shown in Figure 2, falls into this category.

	$\alpha_0$	$\alpha_1$	$\omega_0$	$\omega_1$	$\xi_0$	$\xi_1$	$t_O^*$	$t_M^*$	$t_S^*$	$t_L^*$
$\mathcal{X}_{3G}$	0	0	10	10	100	140	119	125	125	119
$\mathcal{X}_{3LL}$	-4	-4	16	16	112	152	121	128	128	118
$\mathcal{X}_{3RR}$	4	4	16	16	88	128	119	120	123	93
$\mathcal{X}_{3LR}$	-4	4	16	16	112	128	120	124	124	120
$\mathcal{X}_{3RL}$	4	-4	16	16	88	152	119	124	124	123

Table 5: The shape, scale and location parameters for  $\mathcal{X}_{3G}$ ,  $\mathcal{X}_{3LL}$ ,  $\mathcal{X}_{3RR}$ ,  $\mathcal{X}_{3LR}$  and  $\mathcal{X}_{3RL}$ , with corresponding thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$ .

415 In the third set of five data sets,  $\mathcal{X}_{3G}, \dots, \mathcal{X}_{3RL}$ , the sizes of  $\mathcal{C}_0$  and  $\mathcal{C}_1$  are  
 416 very much unbalanced but the variances of them are the same. In this case,  
 417 as shown in Figure 7, all four thresholds are in general well located between  
 418 the two modes of  $\mathcal{C}_0$  and  $\mathcal{C}_1$ , including  $t_O^*$  from Otsu’s method. Since Otsu’s  
 419 method is equivalent to the use of Student’s  $t$ -tests (Xue and Titterington,  
 420 2010b), it can in general be robust to a moderate violation of the assumption  
 421 of normality of the two within-class distributions when the two variances are  
 422 equal.

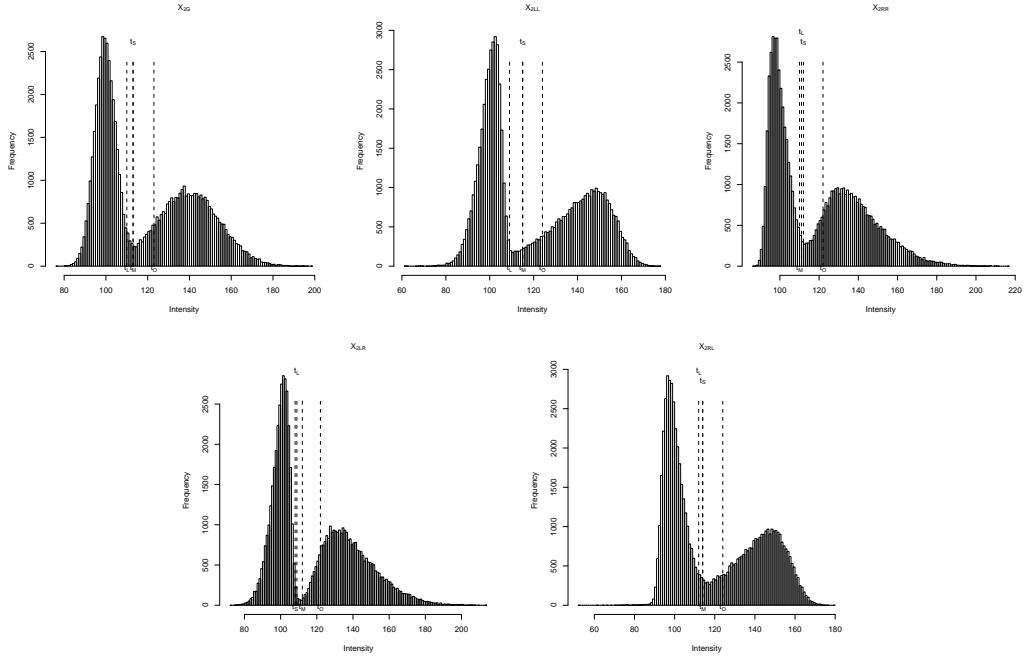


Figure 6: Histograms for  $\mathcal{X}_{2G}$ ,  $\mathcal{X}_{2LL}$ ,  $\mathcal{X}_{2RR}$ ,  $\mathcal{X}_{2LR}$  and  $\mathcal{X}_{2RL}$ , with thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$  indicated by dashed lines.

	$\alpha_0$	$\alpha_1$	$\omega_0$	$\omega_1$	$\xi_0$	$\xi_1$	$t_O^*$	$t_M^*$	$t_S^*$	$t_L^*$
$\mathcal{X}_{4G}$	0	0	5	15	100	140	122	116	115	112
$\mathcal{X}_{4LL}$	-4	-4	8	24	106	159	123	117	112	110
$\mathcal{X}_{4RR}$	4	4	8	24	94	121	121	113	113	112
$\mathcal{X}_{4LR}$	-4	4	8	24	106	121	121	115	108	110
$\mathcal{X}_{4RL}$	4	-4	8	24	94	159	123	116	125	116

Table 6: The shape, scale and location parameters for  $\mathcal{X}_{4G}$ ,  $\mathcal{X}_{4LL}$ ,  $\mathcal{X}_{4RR}$ ,  $\mathcal{X}_{4LR}$  and  $\mathcal{X}_{4RL}$ , with corresponding thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$ .

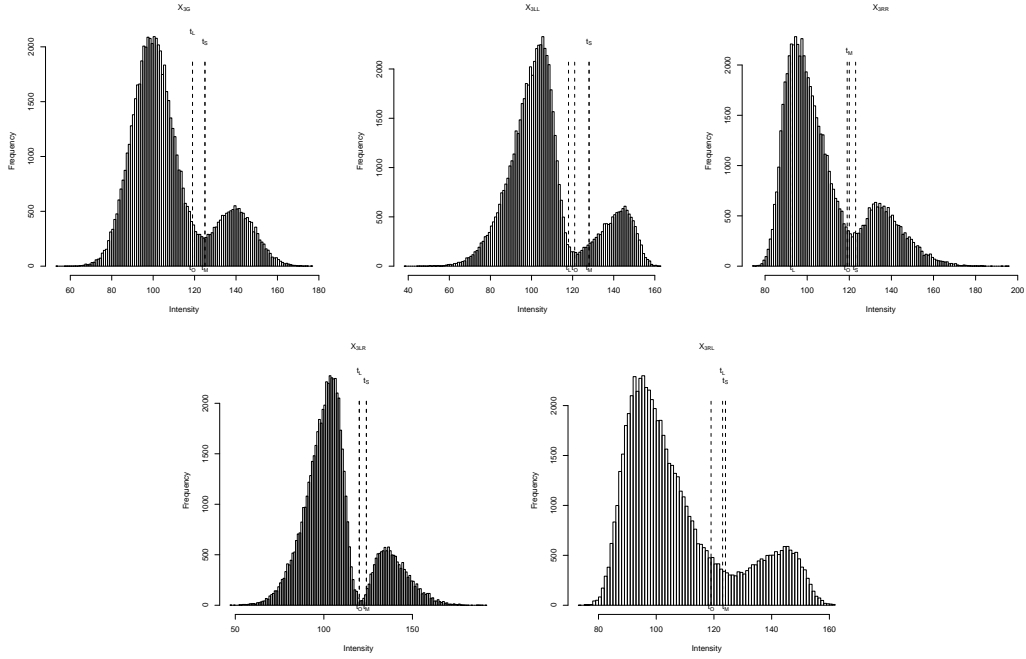


Figure 7: Histograms for  $\mathcal{X}_{3G}$ ,  $\mathcal{X}_{3LL}$ ,  $\mathcal{X}_{3RR}$ ,  $\mathcal{X}_{3LR}$  and  $\mathcal{X}_{3RL}$ , with thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$  indicated by dashed lines.

	$\alpha_0$	$\alpha_1$	$\omega_0$	$\omega_1$	$\xi_0$	$\xi_1$	$t_O^*$	$t_M^*$	$t_S^*$	$t_L^*$
$\mathcal{X}_{5G}$	0	0	15	5	100	140	114	130	130	133
$\mathcal{X}_{5LL}$	-4	-4	24	8	119	146	112	135	135	126
$\mathcal{X}_{5RR}$	4	4	24	8	81	134	115	126	124	134
$\mathcal{X}_{5LR}$	-4	4	24	8	119	134	113	131	131	132
$\mathcal{X}_{5RL}$	4	-4	24	8	81	146	115	128	128	133

Table 7: The shape, scale and location parameters for  $\mathcal{X}_{5G}$ ,  $\mathcal{X}_{5LL}$ ,  $\mathcal{X}_{5RR}$ ,  $\mathcal{X}_{5LR}$  and  $\mathcal{X}_{5RL}$ , with corresponding thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$ .

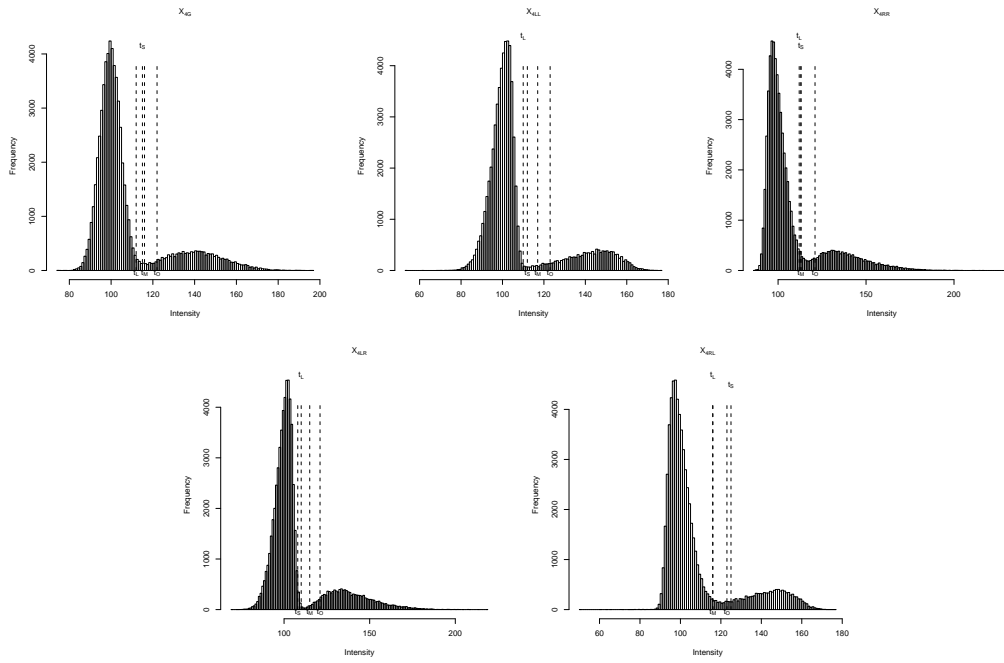


Figure 8: Histograms for  $\mathcal{X}_{AG}$ ,  $\mathcal{X}_{ALL}$ ,  $\mathcal{X}_{ARR}$ ,  $\mathcal{X}_{ALR}$  and  $\mathcal{X}_{ARDL}$ , with thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$  indicated by dashed lines.

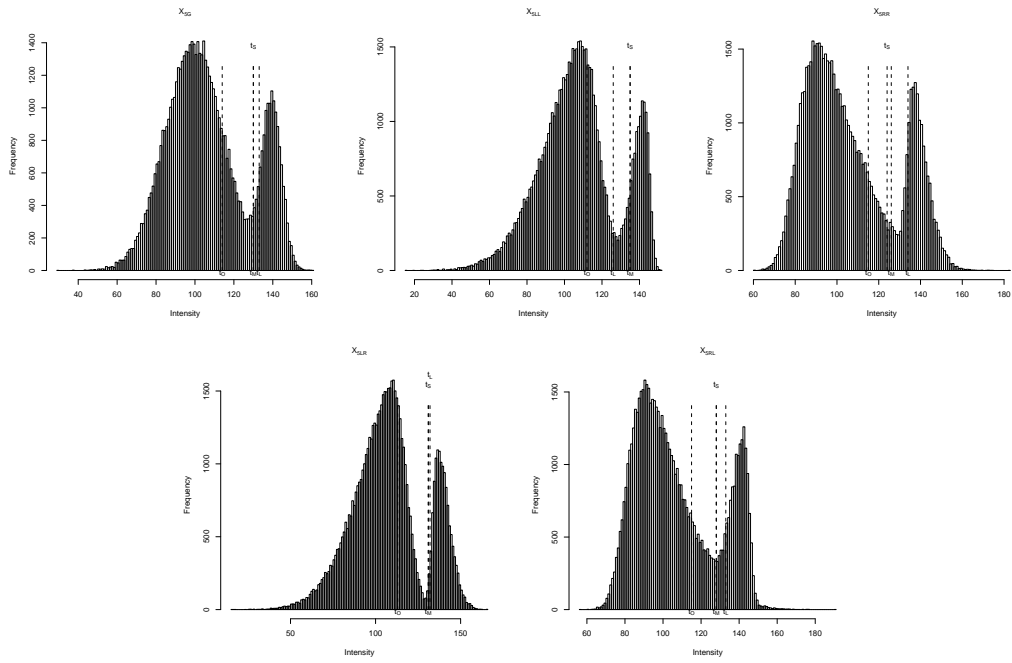


Figure 9: Histograms for  $\mathcal{X}_{5G}$ ,  $\mathcal{X}_{5LL}$ ,  $\mathcal{X}_{5RR}$ ,  $\mathcal{X}_{5LR}$  and  $\mathcal{X}_{5RL}$ , with thresholds  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$  indicated by dashed lines.

423 The fourth set of five data sets,  $\mathcal{X}_{4G}, \dots, \mathcal{X}_{4RL}$ , represent the cases in  
 424 which the sizes of  $\mathcal{C}_0$  and  $\mathcal{C}_1$  are very much unbalanced and the minority class  
 425  $\mathcal{C}_1$  has a larger variance. In contrast, The last five data sets,  $\mathcal{X}_{5G}, \dots, \mathcal{X}_{5RL}$ ,  
 426 represent the cases in which, for two very much unbalanced classes  $\mathcal{C}_0$  and  
 427  $\mathcal{C}_1$ , the majority class  $\mathcal{C}_1$  has a larger variance.

428 As shown in Figures 8 and 9, for the fourth and fifth five categories of  
 429 data, the MET, the skew-normal-based and the log-concave-based methods  
 430 perform similarly well, much better than does Otsu’s method. All thresholds,  
 431  $t_O^*$ ,  $t_M^*$ ,  $t_S^*$  and  $t_L^*$ , are located between the histogram modes for  $\mathcal{C}_0$  and  $\mathcal{C}_1$ ,  
 432 but  $t_O^*$  moves towards the mode of the class with a larger variance.

433 Such patterns are also found in our empirical studies on real images  
 434 ‘Coins’ and ‘NDT-image20’, as the case of ‘Coins’, as shown in Figure 4,  
 435 falls into the fourth category while the case of ‘NDT-image20’, as shown in  
 436 Figure 3, falls into the last category.

437 We also note that, for our simulated data, the skew-normal-based method  
 438 and the MET often provide similar thresholds, implying that, when the skew-  
 439 ness is moderate as in our experiments, the Gaussian mixture is a good ap-  
 440 proximation to the histogram; this may explain to some extent why, in com-  
 441 prehensive studies both for 40 NDT images and for 40 degraded document  
 442 images presented in Sezgin and Sankur (2004), the MET method performs  
 443 the best.

444 In summary, for the binarisation of four real images and twenty-five sim-  
 445 ulated histograms carried out in this paper, our skew-normal-based and log-  
 446 concave-based methods provide results comparable to, if not better than,  
 447 those obtained from the MET method, and provide in general better results  
 448 than that from Otsu’s method.

## 449 6. Discussion

### 450 6.1. Extensions of our skew-normal-based and log-concave-based methods

#### 451 6.1.1. Multi-dimensional extensions

452 Some classical histogram-based image-thresholding methods can provide  
 453 multi-dimensional extensions, for applications in which a multi-dimensional  
 454 (often in practice two-dimensional) histogram, such as that of (grey level,  
 455 locally average grey level) and that of (grey level, grey-level gradient), is  
 456 involved.

457 To extend our one-dimensional skew-normal-based method to such a  
 458 multi-dimensional histogram, the parametric maximum-likelihood estimation

459 of multivariate skew-normal distributions (Azzalini and Valle, 1996; Azzalini  
460 and Capitanio, 1999) can be employed, the implementation of which can be  
461 found in the R package `sn`.

462 Similarly, to achieve a multi-dimensional log-concave-based method, the  
463 nonparametric maximum-likelihood estimation of multivariate log-concave  
464 densities (Cule et al., 2010) should be employed, the implementation of which  
465 can be found in the R package `LogConcDEAD`. We note that, if the dimension  
466 of the histogram is high, then the speed of computation is an issue, and thus  
467 prior to the estimation it is better to perform dimension reduction by using  
468 a technique such as principal component analysis.

### 469 *6.1.2. Multilevel extensions*

470 For a multi-modal histogram, multilevel thresholding can be preferred for  
471 certain applications. In this case, multilevel extensions of the binary skew-  
472 normal-based (or log-concave-based) method proposed in this paper can be  
473 obtained, either by exhaustive search if the number of thresholds is small,  
474 or by using an EM-type algorithm for a finite mixture of skew-normal (or  
475 log-concave) distributions.

476 The EM-type algorithm for a finite skew-normal mixture distribution can  
477 be found in Lin et al. (2007) and Lin (2009), and that for a finite log-concave  
478 mixture distribution can be found in Chang and Walther (2007) and Cule  
479 et al. (2010). Both procedures are naturally analogous to cluster analysis  
480 that is based on finite Gaussian-mixture models (Fraley and Raftery, 2002;  
481 Melnykov and Maitra, 2010).

## 482 *6.2. Limitations of the thresholding methods*

### 483 *6.2.1. Truncation of grey levels*

484 The grey levels are positive numbers, truncated or censored into the fi-  
485 nite range of  $[0, T]$ . Once a threshold  $t$  is fixed, the grey levels smaller than  
486  $t$  are assigned to class  $\mathcal{C}_0$ , and the rest to  $\mathcal{C}_1$ . Hence, the grey levels used to  
487 estimate parameters (such as mean and variance) of  $\mathcal{C}_0$  are truncated into  
488  $[0, t)$ , and those for  $\mathcal{C}_1$  are truncated into  $[t, T]$ . This is a characteristic of  
489 histogram-based thresholding. However, the support of the Gaussian distri-  
490 bution is infinite. This leads to biased estimators of mean and variance for  
491 Otsu's method and the MET method, because both methods provide nonzero  
492 probabilities for the grey levels outside the empirical supports for the object  
493 and the background, but both cannot use these grey levels for estimation.

494 In order to mitigate the effect of truncation on the image-thresholding  
495 methods that assume infinite (and thus overlapped) supports for the two  
496 classes, truncated Gaussian distributions have been used (Lee and Yang,  
497 1989; Cho et al., 1989).

498 It is interesting to investigate the truncation effect on the skew-normal-  
499 based method, since, like the Gaussian distribution, the skew-normal distri-  
500 bution also has an infinite support.

501 For the log-concave-based method, the situation can be even worse. Based  
502 on nonparametric maximum-likelihood estimation, this method assigns zero  
503 probabilities to the grey levels outside the empirical support of  $[0, t)$  for  $\mathcal{C}_0$   
504 and to the grey levels outside  $[t, T]$  for  $\mathcal{C}_1$ , and does not use these grey levels  
505 for estimation. Subsequently, the estimator  $\hat{f}_L(x) = \exp\{\hat{\phi}_c(x)\}$  of  $f_L(x)$   
506 is obtained such that  $\hat{f}_L(x) = 0$  when  $x$  is outside the support  $[x_{(1)}, x_{(N_y)}]$   
507 for class  $\mathcal{C}_y$ . Such a truncation characteristic of the estimator is in general  
508 undesirable, as it provides no probability for unseen extreme observations to  
509 occur.

510 There is a smoothed log-concave density estimator, through the convolu-  
511 tion of  $\hat{f}_L(x)$  and a Gaussian kernel, which can extend the support of  $\hat{f}_L(x)$   
512 to be infinite (Dümbgen and Rufibach, 2010; Cule et al., 2010), but it does  
513 not mitigate the truncation effect for image thresholding.

514 In practice, however, the truncation effect is often disregarded for various  
515 reasons. Its investigation is certainly of research interest, but is beyond the  
516 scope of this paper.

### 517 6.2.2. Model mis-specification

518 The performance of Otsu’s method and the MET method degrades when  
519 the true distributions of the background and the object are far from being  
520 Gaussian distributions. Similarly, the performance of our skew-normal-based  
521 and log-concave-based methods also degrades when the true distributions are  
522 far from being skew-normal or log-concave. Such a mis-specification problem  
523 is associated with all model-based clustering techniques.

524 However, as mentioned in section 2, the skew-normal distribution and the  
525 log-concave distribution include the Gaussian distribution as a special case  
526 and can accommodate some skew distributions. This makes our methods  
527 less sensitive than Otsu’s method or the MET method to the violation of the  
528 assumption of a Gaussian distribution for each class.

529 In addition, when the true distributions  $f(x)$  of the background and the  
530 object are not log-concave, the nonparametric maximum-likelihood estima-

531 tors  $\hat{f}_L(x)$  of  $f(x)$  may converge in certain norms to the log-concave densities  
532 that are closest to  $f(x)$  in some sense (Cule et al., 2010).

### 533 6.3. Evaluation of image-thresholding methods

534 To evaluate and compare image-binarisation methods, a considerable  
535 number of measures have been proposed (Sahoo et al., 1988; Sezgin and  
536 Sankur, 2004; Zhang, 1996; Zhang et al., 2008). For real images, because in  
537 general no pre-segmented image is available as ‘ground truth’, objective eval-  
538 uation criteria, such as uniformity and shape measures, are often adopted.

539 However, such ‘objective’ evaluation prefers a method for which the rule  
540 for selecting  $t^*$  matches the evaluation criteria. For example, the grey-level  
541 uniformity measure proposed by Levine and Nazif (1985) was used in com-  
542 prehensive surveys and comparative studies such as Sahoo et al. (1988), Lee  
543 et al. (1990), Zhang (1996) and Sezgin and Sankur (2004), among others.  
544 However, this measure is essentially equivalent to Otsu’s rule for selecting  
545 an optimal threshold (Ng and Lee, 1996) and thus always favours Otsu’s  
546 method. Therefore, we opt not to use these quantitative measures to evaluate  
547 the image-thresholding methods that are compared in this paper. Moreover,  
548 for the simulated histograms and the real images shown in section 5, visual  
549 perception provides reasonable comparison among the image-thresholding  
550 methods, although there is ‘ground truth’ kindly provided by Sezgin and  
551 Sankur (2004) for the NDT images.

## 552 7. Conclusions

553 Two new methods for histogram-based image thresholding have been  
554 proposed: one is based on parametric maximum-likelihood estimation of  
555 skew-normal distributions; the other is based on nonparametric maximum-  
556 likelihood estimation of log-concave distributions. The two methods gener-  
557 alised the classical Gaussian-based methods, considered the skewness of the  
558 distributions of individual classes, and based the determination of an optimal  
559 threshold on the comparison of maximum log-likelihoods. The experiments  
560 presented in this paper, which applied the methods to real images and simu-  
561 lated data, demonstrated that the two methods performed as well as, if not  
562 better than, Otsu’s method and the MET method, two of the most popular  
563 approaches to image thresholding.

564 **Acknowledgments**

565 The NDT images were made available through the courtesy of Dr. Mehmet  
566 Sezgin. This work was partly supported by funding to J.-H.X. from the  
567 Internal Visiting Programme, under the EU-funded PASCAL2 Network of  
568 Excellence.

569 **References**

- 570 Azzalini, A., 1985. A class of distributions which includes the normal ones.  
571 *Scandinavian Journal of Statistics* 12 (2), 171–178.
- 572 Azzalini, A., Capitanio, A., 1999. Statistical applications of the multivariate  
573 skew normal distribution. *Journal of the Royal Statistical Society, Series*  
574 *B* 61 (3), 579–602.
- 575 Azzalini, A., Valle, A. D., 1996. The multivariate skew-normal distribution.  
576 *Biometrika* 83 (4), 715–726.
- 577 Bazi, Y., Bruzzone, L., Melgani, F., 2007. Image thresholding based on the  
578 EM algorithm and the generalized Gaussian distribution. *Pattern Recog-*  
579 *nition* 40 (2), 619–634.
- 580 Chang, G. T., Walther, G., 2007. Clustering with mixtures of log-concave  
581 distributions. *Computational Statistics & Data Analysis* 51 (12), 6242–  
582 6251.
- 583 Cho, S., Haralick, R., Yi, S., 1989. Improvement of Kittler and Illingworth’s  
584 minimum error thresholding. *Pattern Recognition* 22 (5), 609–617.
- 585 Cule, M., Gramacy, R., Samworth, R., 2009. LogConcDEAD: An R package  
586 for maximum likelihood estimation of a multivariate log-concave density.  
587 *Journal of Statistical Software* 29 (2).  
588 URL <http://www.jstatsoft.org/v29/i02/>
- 589 Cule, M., Samworth, R., Stewart, M., 2010. Maximum likelihood estimation  
590 of a multi-dimensional log-concave density (with discussion). *Journal of*  
591 *the Royal Statistical Society, Series B* 72 (5), to appear.
- 592 Dümbgen, L., Hüsler, A., Rufibach, K., 2007. Active set and EM algo-  
593 rithms for log-concave densities based on complete and censored data.  
594 Tech. Rep. 61, Universität Bern, Bern.

- 595 Dümbgen, L., Rufibach, K., 2010. logcondens: Computations related to  
596 univariate log-concave density estimation. *Journal of Statistical Software*  
597 30 (to appear).
- 598 Fan, S.-K. S., Lin, Y., 2009. A fast estimation method for the generalized  
599 Gaussian mixture distribution on complex images. *Computer Vision and*  
600 *Image Understanding* 113 (7), 839–853.
- 601 Fan, S.-K. S., Lin, Y., Wu, C.-C., 2008. Image thresholding using a novel  
602 estimation method in generalized Gaussian distribution mixture modeling.  
603 *Neurocomputing* 72 (1-3), 500–512.
- 604 Fraley, C. F., Raftery, A. E., 2002. Model-based clustering, discriminant  
605 analysis, and density estimation. *Journal of the American Statistical As-*  
606 *sociation* 97 (458), 611–631.
- 607 Glasbey, C. A., 1993. An analysis of histogram-based thresholding algo-  
608 rithms. *CVGIP: Graphical Models and Image Processing* 55 (6), 532–537.
- 609 Gupta, R. C., Balakrishnan, N., 2010. Log-concavity and monotonicity of  
610 hazard and reversed hazard functions of univariate and multivariate skew-  
611 normal distributions. *Metrika* DOI: 10.1007/s00184-010-0321-9.
- 612 Kittler, J., Illingworth, J., 1986. Minimum error thresholding. *Pattern Recog-*  
613 *nition* 19 (1), 41–47.
- 614 Kurita, T., Otsu, N., Abdelmalek, N., 1992. Maximum likelihood thresh-  
615 olding based on population mixture models. *Pattern Recognition* 25 (10),  
616 1231–1240.
- 617 Lee, J. S., Yang, M. C. K., 1989. Threshold selection using estimates from  
618 truncated normal distribution. *IEEE Transactions on Systems, Man and*  
619 *Cybernetics* 19 (2), 422–429.
- 620 Lee, S. U., Chung, S. Y., Park, R. H., 1990. A comparative performance  
621 study of several global thresholding techniques for segmentation. *Computer*  
622 *Vision, Graphics, and Image Processing* 52 (2), 171–190.
- 623 Levine, M. D., Nazif, A. M., 1985. Dynamic measurement of computer gen-  
624 erated image segmentations. *IEEE Transactions on Pattern Analysis and*  
625 *Machine Intelligence* 7 (2), 155–164.

- 626 Lin, T. I., 2009. Maximum likelihood estimation for multivariate skew normal  
627 mixture models. *Journal of Multivariate Analysis* 100 (2), 257–265.
- 628 Lin, T. I., Lee, J. C., Yen, S. Y., 2007. Finite mixture modelling using the  
629 skew normal distribution. *Statistica Sinica* 17 (3), 909–927.
- 630 Melnykov, V., Maitra, R., 2010. Finite mixture models and model-based  
631 clustering. *Statistics Surveys* 4, 80–116.
- 632 Moser, G., Serpico, S. B., 2006. Generalized minimum-error thresholding  
633 for unsupervised change detection from SAR amplitude imagery. *IEEE*  
634 *Transactions on Geoscience and Remote Sensing* 44 (10), 2972–2982.
- 635 Ng, W. S., Lee, C. K., 1996. Comment on using the uniformity measure  
636 for performance measure in image segmentation. *IEEE Transactions on*  
637 *Pattern Analysis and Machine Intelligence* 18 (9), 933–934.
- 638 Otsu, N., 1979. A threshold selection method from gray-level histograms.  
639 *IEEE Transactions on Systems, Man, and Cybernetics* SMC-9, 62–66.
- 640 Pal, N. R., Bhandari, D., 1993. Image thresholding: some new techniques.  
641 *Signal Processing* 33 (2), 139–158.
- 642 Pal, N. R., Pal, S. K., 1993. A review on image segmentation techniques.  
643 *Pattern Recognition* 26 (9), 1277–1294.
- 644 Rufibach, K., 2007. Computing maximum likelihood estimators of a log-  
645 concave density function. *Journal of Statistical Computation and Simula-*  
646 *tion* 77 (7), 561–574.
- 647 Sahoo, P. K., Soltani, S., Wong, A. K. C., Chen, Y. C., 1988. A survey of  
648 thresholding techniques. *Computer Vision, Graphics, and Image Process-*  
649 *ing* 41 (2), 233–260.
- 650 Sezgin, M., Sankur, B., 2004. Survey over image thresholding techniques and  
651 quantitative performance evaluation. *Journal of Electronic Imaging* 13 (1),  
652 146–165.
- 653 Trier, Ø. D., Jain, A. K., 1995. Goal-directed evaluation of binarization  
654 methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*  
655 17 (12), 1191–1201.

- 656 Walther, G., 2002. Detecting the presence of mixing with multiscale maxi-  
657 mum likelihood. *Journal of the American Statistical Association* 97 (458),  
658 508–513.
- 659 Walther, G., 2009. Inference and modeling with log-concave distributions.  
660 *Statistical Science* 24 (3), 319–327.
- 661 Xue, J.-H., Titterington, D. M., 2010a. Image thresholding by using statistics  
662 for determining the number of clusters. manuscript.
- 663 Xue, J.-H., Titterington, D. M., 2010b.  $t$ -tests,  $F$ -tests and Otsu’s methods  
664 for image thresholding. manuscript.
- 665 Xue, J.-H., Zhang, Y. J., Lin, X. G., 1999. Rayleigh-distribution based min-  
666 imum error thresholding for SAR images. *Journal of Electronics (China)*  
667 16 (4), 336–342.
- 668 Yan, H., 1996. Unified formulation of a class of image thresholding techniques.  
669 *Pattern Recognition* 29 (12), 2025–2032.
- 670 Zhang, H., Fritts, J. E., Goldman, S. A., 2008. Image segmentation eval-  
671 uation: a survey of unsupervised methods. *Computer Vision and Image*  
672 *Understanding* 110 (2), 260–280.
- 673 Zhang, Y. J., 1996. A survey on evaluation methods for image segmentation.  
674 *Pattern Recognition* 29 (8), 1335–1346.