

# Median-based Image Thresholding

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

In order to select for image thresholding an optimal threshold that is relatively robust to the presence of skew or heavy-tailed class-conditional distributions, we propose in this paper two median-based approaches: one is an extension of Otsu's method, and the other is an extension of Kittler and Illingworth's minimum error thresholding. The two extensions preserve the methodological simplicity and computational efficiency of their original methods. Experiments on some real images and simulated data sets show that the two extensions can accomplish robust performance. In addition, theoretical interpretation of the new approaches, based on the mixture of Laplace distributions, is given.

*Keywords:* Image segmentation, Image thresholding, Laplace distributions, Mean absolute deviation from the median (MAD), Minimum error thresholding (MET), Otsu's method.

---

## 1. Introduction

Image thresholding aims to partition an image into  $K$  predetermined, mutually-exclusive classes,  $C_1, \dots, C_K$ , based on  $K - 1$  grey-level thresholds. Most commonly,  $K = 2$  such that the image is partitioned into the background and the foreground. As an initial procedure for realising image segmentation, it has a long history of investigation, motivated by a broad range of practical applications of image analysis and object recognition. Comprehensive overviews and comparative studies of image thresholding can be

---

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

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

9 found in Sahoo et al. (1988), Glasbey (1993), Trier and Jain (1995) and  
10 Sezgin and Sankur (2004), for example.

11 Many, and the most-widely used, approaches to image thresholding are  
12 based on analysis of the histogram of grey levels (or termed intensities) in an  
13 image, searching for an optimal grey-level threshold  $t^*$  to divide the histogram  
14 into two parts,  $C_1$  with grey levels lower than  $t^*$  and  $C_2$  for the upper part.

15 Among these approaches, two of the most popular are Otsu's method  
16 (Otsu, 1979) and Kittler and Illingworth's minimum-error-thresholding (MET)  
17 method (Kittler and Illingworth, 1986). Otsu's method is adopted as the  
18 method for automatic image thresholding by some free and commercial soft-  
19 ware, such as GIMP ([www.gimp.org](http://www.gimp.org)) and MATLAB (The MathWorks, Inc.).  
20 The MET method ranks the best in a comprehensive survey of image thresh-  
21 olding conducted by Sezgin and Sankur (2004).

22 In image thresholding, an optimal threshold  $t^*$  is often determined based  
23 on the estimation of locations and dispersions of  $C_1$  and  $C_2$ . As with many  
24 other approaches, both Otsu's method and the MET method use the sample  
25 mean and the sample standard deviation to estimate location and dispersion,  
26 respectively.

27 It is well-known that, when the distribution for class  $C_k$  is skew or heavy-  
28 tailed, or when there are outliers in the sample from  $C_k$ , the median is a more  
29 robust estimator of location than the mean. When the median is chosen for  
30 location, the mean absolute deviation from the median (denoted by MAD  
31 hereafter) is usually chosen as the estimator of dispersion.

32 Therefore, in order to select a  $t^*$  that is more robust to the presence  
33 of skew or heavy-tailed distributions for  $C_k$  than those selected by Otsu's  
34 method and the MET method, we propose in section 2 two median-based  
35 approaches to image thresholding. One of them is an extension of Otsu's  
36 method and the other is an extension of the MET method, both based on  
37 the use of the MAD. Meanwhile, like their original versions, the two new  
38 approaches remain methodologically simple and computationally efficient.

39 In addition, the relationship between Otsu's method and the MET method  
40 has been investigated by Kurita et al. (1992) and Yan (1996). Kurita et al.  
41 (1992) show that both methods can be derived from maximisation of the  
42 log-likelihoods based on mixtures of Gaussian distributions. In section 3, we  
43 present a theoretical interpretation of the median-based methods from the  
44 perspective of maximisation of the log-likelihoods for mixtures of Laplace  
45 distributions.

46 For illustrative and comparative purposes, experiments of applying the

47 median-based methods and their original versions to some real images and  
 48 some histograms constructed from simulated data are presented in section 4,  
 49 and some conclusions are drawn in section 5.

## 50 2. Methodology

51 For an image  $\mathcal{X}$  of  $N$  pixels, each pixel is represented by its grey level  
 52  $x_i, i = 1, \dots, N$ . A grey-level threshold  $t$  partitions the image into two classes  
 53  $C_1(t)$  and  $C_2(t)$ , where  $C_1(t) = \{i : 0 \leq x_i \leq t, 1 \leq i \leq N\}$  and  $C_2(t) = \{i :$   
 54  $t < x_i \leq T, 1 \leq i \leq N\}$ , in which  $T$  is the largest possible grey level, which  
 55 is 255 for an 8-bit grey-level image (i.e.  $x_i \in [0, T]$ ).

56 The histogram for the image  $\mathcal{X}$ , denoted by  $h(x)$  hereafter, can be con-  
 57 structed by counting the frequencies of the grey levels and dividing them by  
 58  $N$ , such that  $\sum_{x=0}^T h(x) = 1$ .

### 59 2.1. Otsu's method and its median-based extension

#### 60 2.1.1. Otsu's method

61 Otsu's rule (Otsu, 1979; Kurita et al., 1992) for selecting the optimal  
 62 threshold  $t^*$  can be written as

$$t_O^* = \underset{t}{\operatorname{argmin}} \left\{ \omega_1(t)s_1^2(t) + \omega_2(t)s_2^2(t) \right\} , \quad (1)$$

63 where  $\omega_1(t)$  and  $\omega_2(t)$  are the proportions of pixels representing classes  $C_1(t)$   
 64 and  $C_2(t)$  determined by a threshold  $t$ ,  $s_1(t)$  and  $s_2(t)$  are the (biased) sample  
 65 standard deviations for  $C_1(t)$  and  $C_2(t)$ , respectively, defined as

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

66

$$s_1(t) = \sum_{x=0}^t \left\{ \frac{h(x)}{\omega_1(t)} (x - \bar{x}_1)^2 \right\} , \quad s_2(t) = \sum_{x=t+1}^T \left\{ \frac{h(x)}{\omega_2(t)} (x - \bar{x}_2)^2 \right\} , \quad (3)$$

67 in which  $\bar{x}_1(t) = \sum_{x=0}^t \{xh(x)/\omega_1(t)\}$  and  $\bar{x}_2(t) = \sum_{x=t+1}^T \{xh(x)/\omega_2(t)\}$  are  
 68 the sample means for  $C_1(t)$  and  $C_2(t)$ , respectively.

69 *2.1.2. A median-based extension*

70 As mentioned in section 1, we envision that the use of the median instead  
 71 of the mean may provide a  $t^*$  that is more robust to the presence of skew or  
 72 heavy-tailed distributions for  $C_k$  than those selected by Otsu's method and  
 73 the MET method. Therefore, a median-based extension of Otsu's method,  
 74 naturally derived from substituting the MAD for  $s^2$  (not for  $s$  here for theo-  
 75 retical reasons explained in section 3), provides a rule for selecting  $t^*$  (denoted  
 76 by  $g^*$  for distinctive purposes hereafter) as follows:

$$g_O^* = \operatorname{argmin}_t \{ \omega_1(t) \text{MAD}_1(t) + \omega_2(t) \text{MAD}_2(t) \} , \quad (4)$$

77 where  $\text{MAD}_k(t)$ , the mean absolute deviations from the median for class  
 78  $C_k(t)$ , are given by

$$\text{MAD}_1(t) = \sum_{x=0}^t \left\{ \frac{h(x)}{\omega_1(t)} |x - m_1(t)| \right\} , \quad (5)$$

$$\text{MAD}_2(t) = \sum_{x=t+1}^T \left\{ \frac{h(x)}{\omega_2(t)} |x - m_2(t)| \right\} , \quad (6)$$

79 in which  $m_1(t) = \text{med}\{x_i : i \in C_1(t)\}$  and  $m_2(t) = \text{med}\{x_i : i \in C_2(t)\}$  are  
 80 the sample medians for  $C_1(t)$  and  $C_2(t)$ , respectively.

81 *2.1.3. Multi-level thresholding*

82 When there are more than two classes predetermined for an image (i.e.  
 83  $K > 2$ ), it would be better to use more-than-one thresholds to partition the  
 84 image into these classes, leading to a multi-level thresholding problem.

85 For multi-level thresholding, Otsu's rule for selecting optimal thresholds  
 86  $\mathbf{t}^* = (t_1^*, \dots, t_{K-1}^*)$  can be written as

$$\mathbf{t}_O^* = \operatorname{argmin}_{\mathbf{t}} \sum_{k=1}^K \{ \omega_k(\mathbf{t}) s_k^2(\mathbf{t}) \} , \quad (7)$$

87 where, similarly to that in section 2.1.1,  $\omega_k(t)$  and  $s_k^2(t)$  are defined for  $C_k(t)$ .

88 Therefore, for multi-level thresholding, the rule of Otsu's median-based  
 89 extension becomes

$$\mathbf{g}_O^* = \operatorname{argmin}_{\mathbf{t}} \sum_{k=1}^K \{ \omega_k(\mathbf{t}) \text{MAD}_k(\mathbf{t}) \} . \quad (8)$$

90 *2.2. The MET method and its median-based extension*

91 *2.2.1. The MET method*

92 The MET method (Kittler and Illingworth, 1986) selects  $t^*$  as

$$t_M^* = \operatorname{argmin}_t \left\{ \omega_1(t) \log \frac{s_1(t)}{\omega_1(t)} + \omega_2(t) \log \frac{s_2(t)}{\omega_2(t)} \right\}, \quad (9)$$

93 where  $\omega_1(t)$ ,  $\omega_2(t)$ ,  $s_1(t)$  and  $s_2(t)$ , defined in equations (2) and (3), are  
94 positive here.

95 *2.2.2. A median-based extension*

96 In analogy to that in section 2.1, the rule of a median-based extension of  
97 the MET method can be derived from substituting the MAD for  $s$  (not for  
98  $s^2$  here for theoretical reasons explained in section 3) as

$$g_M^* = \operatorname{argmin}_t \left\{ \omega_1(t) \log \frac{\text{MAD}_1(t)}{\omega_1(t)} + \omega_2(t) \log \frac{\text{MAD}_2(t)}{\omega_2(t)} \right\}. \quad (10)$$

99 *2.2.3. Multi-level thresholding*

100 The multi-level-thresholding versions of the MET method and its median-  
101 based extension are readily written as

$$\mathbf{t}_M^* = \operatorname{argmin}_{\mathbf{t}} \sum_{k=1}^K \left\{ \omega_k(\mathbf{t}) \log \frac{s_k(\mathbf{t})}{\omega_k(\mathbf{t})} \right\}, \quad (11)$$

102

$$\mathbf{g}_M^* = \operatorname{argmin}_{\mathbf{t}} \sum_{k=1}^K \left\{ \omega_k(\mathbf{t}) \log \frac{\text{MAD}_k(\mathbf{t})}{\omega_k(\mathbf{t})} \right\}. \quad (12)$$

### 103 **3. Theoretical interpretation**

104 A straightforward and intuitive interpretation of Otsu's rule, as shown  
105 in equation (1), is that it aims to minimise the within-classes variance, a  
106 measure of dispersion, of the grey level, which is represented by the term to  
107 be minimised in the right-hand side of equation (1). Correspondingly, the  
108 interpretation of the median-based extension of Otsu's method, as shown in  
109 equation (4), can be that the extension aims to minimise the within-classes  
110 mean absolute deviation from the median, another measure of dispersion, of  
111 the grey level.

112 Alternatively and insightfully, as mentioned in section 1, Kurita et al.  
 113 (1992) show that both Otsu’s method and the MET method can be derived  
 114 from maximisation of the log-likelihoods based on mixtures of Gaussian dis-  
 115 tributions. The same spirit can be found in Kittler and Illingworth (1986)  
 116 from the derivation of the MET method, although not explicitly mentioned.  
 117 Analogous to that, here we also present such a theoretical interpretation of  
 118 the MAD-based methods, from the perspective of maximisation of the log-  
 119 likelihoods but based on mixtures of Laplace distributions, as follows.

120 Suppose that the grey level of class  $C_k$  follows a Laplace distribution, of  
 121 which the probability density function  $p(x|C_k)$  is defined as

$$p(x|C_k) = \frac{1}{2\beta_k} \exp\left(-\frac{|x - \alpha_k|}{\beta_k}\right), \quad (13)$$

122 where  $\alpha_k$  is a location parameter and  $\beta_k$  is a positive scale parameter. The  
 123 maximum likelihood estimator (MLE) of  $\alpha_k$  is the sample median  $m_k$  for  $C_k$ ,  
 124 and the MLE of  $\beta_k$  is  $\text{MAD}_k$ , the mean absolute deviation from the median  
 125 for  $C_k$ .

126 Without loss of generality, here we only consider the case of image bina-  
 127 risation (i.e. into  $C_1$  and  $C_2$ ).

128 In this case, the log-likelihood based on the joint distribution of the grey  
 129 level and its class indicator can be written as

$$\begin{aligned} \ell(\theta(t)) = & \sum_{x=0}^t \left[ h(x)N \left\{ \log(p_1(t)) - \log(2\beta_1(t)) - \frac{|x - \alpha_1(t)|}{\beta_1(t)} \right\} \right] \\ & + \sum_{x=t+1}^T \left[ h(x)N \left\{ \log(p_2(t)) - \log(2\beta_2(t)) - \frac{|x - \alpha_2(t)|}{\beta_2(t)} \right\} \right], \end{aligned} \quad (14)$$

130 where  $\theta(t) = \{\alpha_1(t), \beta_1(t), \alpha_2(t), \beta_2(t)\}$ , and  $p_k(t)$  is the prior probability of  
 131 class  $C_k(t)$ , of which the MLE is the proportion  $\omega_k(t)$ .

132 As we mentioned, the MLE of  $p_k(t)$ ,  $\alpha_k(t)$  and  $\beta_k(t)$  are  $\omega_k(t)$ ,  $m_k(t)$  and  
 133  $\text{MAD}_k(t)$ , respectively, as shown in equations (2), (5) and (6). Plugging them  
 134 into equation (14), we get

$$\begin{aligned} \max_{\theta(t)} \frac{\ell(\theta(t))}{N} &= \omega_1(t) \log(\omega_1(t)) - \omega_1(t) \log(2\text{MAD}_1(t)) - \omega_1(t) \\ &\quad + \omega_2(t) \log(\omega_2(t)) - \omega_2(t) \log(2\text{MAD}_2(t)) - \omega_2(t), \\ &= \omega_1(t) \log \frac{\omega_1(t)}{\text{MAD}_1(t)} + \omega_2(t) \log \frac{\omega_2(t)}{\text{MAD}_2(t)} - 1 - \log 2. \end{aligned} \quad (15)$$

135 Therefore, we can select the optimal  $g^*$  that provides the largest maximum  
 136 log-likelihood as follows:

$$g^* = \operatorname{argmax}_t \left\{ \omega_1(t) \log \frac{\omega_1(t)}{\operatorname{MAD}_1(t)} + \omega_2(t) \log \frac{\omega_2(t)}{\operatorname{MAD}_2(t)} \right\}, \quad (16)$$

137 which equals  $g_M^*$  in equation (10), the threshold selected by the MAD-based  
 138 extension of the MET method. This provides an interpretation for such a  
 139 MAD-based extension.

140 When we consider the log-likelihood based only on the class-conditional  
 141 distribution  $p(x|C_k)$ , in equation (14) the terms related to  $\log(p_k(t))$  dis-  
 142 appear. We further assume that  $\beta_1(t) = \beta_2(t) = \beta_w(t)$  such that  $\beta_w(t)$  is  
 143 estimated by the within-classes mean absolute deviation from the median,  
 144 which is calculated as  $\omega_1(t)\operatorname{MAD}_1(t) + \omega_2(t)\operatorname{MAD}_2(t)$ . Consequently, based  
 145 on a derivation in the same manner as that from equation (14) to equation  
 146 (16), we can readily obtain the MAD-based extension of Otsu’s method, as  
 147 shown in equation (4).

148 This also explains why we substitute the MAD for  $s^2$  when we derive  
 149 the extension of Otsu’s method, while substituting the MAD for  $s$  when  
 150 we derive the extension of the MET method. Nonetheless, we note that,  
 151 given robust thresholding performance of the MAD-based methods, other  
 152 theoretical justification of them merit further investigation.

153 Besides the methods that model the histogram as a mixture of Gaus-  
 154 sian distributions (Kittler and Illingworth, 1986; Kurita et al., 1992), there  
 155 are other image-thresholding methods, which model the histogram as a mix-  
 156 ture of Poisson (Pal and Bhandari, 1993), generalised Gaussian distributions  
 157 (Bazi et al., 2007; Fan et al., 2008), skew-normal and log-concave distribu-  
 158 tions (Xue and Titterington, 2010c) and certain distributions derived from  
 159 Rayleigh (Xue et al., 1999), Nakagami-gamma, log-normal and Weibull dis-  
 160 tributions (Moser and Serpico, 2006). Some of these distributions, such as  
 161 the generalised Gaussian distribution and the log-concave distribution, in-  
 162 clude the Laplace distribution as a special case. However, fitting a mixture  
 163 of such (relatively generic) distributions is complicated in both theoretical  
 164 and computational senses.

## 165 4. Experimental Studies

166 For illustrative and comparative purposes, we present the results obtained  
 167 from applying the four methods (Otsu’s method, the MET method and their

168 median-based extensions), first to fifteen simulated data sets in section 4.1  
169 and then to six real images in section 4.2.

#### 170 4.1. Simulated data sets

##### 171 4.1.1. Data

172 When the class-conditional distribution of the intensity for each class  
173  $C_k$  is close to a Gaussian distribution, the Gaussian-mixture assumption  
174 underlying the MET method or Otsu’s method, as mentioned in section 3,  
175 can be approximately satisfied. Hence, in this case, the MET method and  
176 Otsu’s method can in general be recommended.

177 The purpose for proposing in this paper the two median-based extensions  
178 of Otsu’s method and the MET method is to find a more robust thresholding  
179 approach, which is less sensitive to the presence of heavy-tailed or skew class-  
180 conditional distributions, than the original Otsu and MET methods.

181 In this context, in order to demonstrate some scenarios that the two  
182 median-based approaches can be recommended for, we simulate intensities  
183 from two-component mixtures of skew  $t$ -distributions (Azzalini and Capi-  
184 tano, 2003). A skew  $t$ -distribution is skew and heavy-tailed.

185 With such a mixture, the factors influencing the thresholding performance  
186 include the degrees of skewness and leptokurtosis of each class-conditional  
187 skew  $t$ -distribution, the directions of the skewness, the ratio between two  
188 class sizes and the ratio between two within-classes variances.

189 Among these factors, the two ratios are important to Otsu’s method  
190 and its median-based extension, because for these two methods equal sizes  
191 and a common variance (or MAD) are assumed between the two classes, as  
192 discussed in section 3.

193 In the case that the sample sizes are roughly equal, the influence of un-  
194 equal variances on Otsu’s method is not remarkable (Xue and Titterington,  
195 2010b), as Otsu’s method can be regarded as using Student’s  $t$ -tests. In  
196 addition, our experiments with simulated data from class-conditional skew  
197  $t$ -distributions suggest that the performances of all four methods studied in  
198 this paper are also comparable in this case. Therefore, here we only consider  
199 the cases in which the sizes of the two classes ( $C_1$  and  $C_2$ ) are relatively  
200 distinct (in a ratio of 4 : 1) and, without loss of generality, the class  $C_1$  for  
201 the lower part of the histogram is the majority class ( $\pi_1 = 80\%$ ), and the  
202 upper-part class  $C_2$  is the minority class. The class proportions  $\pi_k$  and the  
203 standard deviations  $\sigma_k$  of  $C_k$  are listed in Table 1, in which there are three

204 unique relationships between  $\sigma_1$  and  $\sigma_2$ :  $\sigma_1 = \sigma_2$ ,  $\sigma_1 < \sigma_2$  and  $\sigma_1 > \sigma_2$ ,  
 205 denoted by  $\mathcal{X}_a$ ,  $\mathcal{X}_b$  and  $\mathcal{X}_c$ , respectively.

	$\pi_1$	$\pi_2$	$\sigma_1$	$\sigma_2$
$\mathcal{X}_a$	0.8	0.2	10	10
$\mathcal{X}_b$	0.8	0.2	5	15
$\mathcal{X}_c$	0.8	0.2	15	5

Table 1: Three unique relationships between standard deviations  $\sigma_1$  and  $\sigma_2$  and the values of proportions  $\pi_1$  and  $\pi_2$  used for data simulation in this paper.

206 A skew  $t$ -distribution can be denoted by  $\text{St}(\xi, \omega, \alpha, \nu)$ , where  $\xi$ ,  $\omega$ ,  $\alpha$   
 207 and  $\mu$  represent the parameters for the location, scale, shape and degrees of  
 208 freedom, respectively. The values of these parameters are set as follows for  
 209 data simulation.

210 First, the degrees of freedom  $\nu$  for both class-conditional distributions are  
 211 set to 3, indicating heavy-tailed (or leptokurtic) distributions.

212 Secondly, the shape parameter  $\alpha$  regulates the skewness of the distribu-  
 213 tion, with its absolute value for the degree of the skewness and its sign for  
 214 the direction of the skewness. In our studies,  $\alpha = 0$  for a symmetric class-  
 215 conditional distribution,  $\alpha = 4$  for a right-skewed (or positively-skewed) dis-  
 216 tribution and  $\alpha = -4$  for a left-skewed (or negatively-skewed) distribution.

217 For each of the three relationships in Table 1, we consider five types of  
 218 skewness, denoted by  $S$ ,  $LR$ ,  $LL$ ,  $RR$  and  $RL$ , respectively. Of the five types,  
 219  $S$  corresponds to a scenario that both class-conditional distributions are sym-  
 220 metric;  $LR$  represents a scenario that the distribution for  $C_1$  is left-skewed  
 221 (with  $L$  standing for left-skewed) and the distribution for  $C_2$  is right-skewed  
 222 (with  $R$  for right-skewed), indicating that the two class-conditional distribu-  
 223 tions skew away from each other;  $LL$  is for a scenario that the minority class  
 224  $C_2$  skews towards the majority class  $C_1$ ;  $RR$  points to a scenario that the  
 225 majority class skews towards the minority class; and  $RL$  presents a scenario  
 226 that the two distributions skew towards each other. Therefore, in total, we  
 227 build from simulated intensities fifteen data sets, denoted by  $\mathcal{X}_{aS}, \dots, \mathcal{X}_{cRL}$ ,  
 228 respectively, for all combinations of the three relationships between variances  
 229 and the five types of skewness.

230 Thirdly, in order to make the results for all fifteen data sets and the  
 231 corresponding conclusions compatible, we calibrate the values of the scale  
 232 parameter  $\omega$  and the location parameter  $\xi$  for each data set, such that for  
 233 every data set it is approximately satisfied that  $\mu_1 = 100$ ,  $\mu_2 = 140$ , and  $\sigma_1$

234 and  $\sigma_2$  equal the corresponding values in Table 1. The calibration can be  
 235 achieved based on the property that the mean  $\mu$  and the variance  $\sigma^2$  of a  
 236 skew  $t$ -distribution  $\text{St}(\xi, \omega, \alpha, \nu)$  can be given by

$$\mu = \xi + \omega\eta, \quad (17)$$

$$\sigma^2 = \omega^2 \left( \frac{\nu}{\nu-2} - \eta^2 \right), \quad (18)$$

237 where  $\eta = \delta\Gamma\{(\nu-1)/2\}/\Gamma(\nu/2)(\nu/\pi)^{1/2}$ , in which  $\delta = \alpha/\sqrt{1+\alpha^2}$  and  $\pi$  is  
 238 Archimedes' constant (Azzalini and Capitanio, 2003).

239 With all parameters set, we adopt an R package `sn` to simulate data from  
 240 the  $\text{St}(\xi_k, \omega_k, \alpha_k, \nu_k)$  distribution for class  $C_k$ , for  $k = 1$  and 2.

#### 241 4.1.2. Results

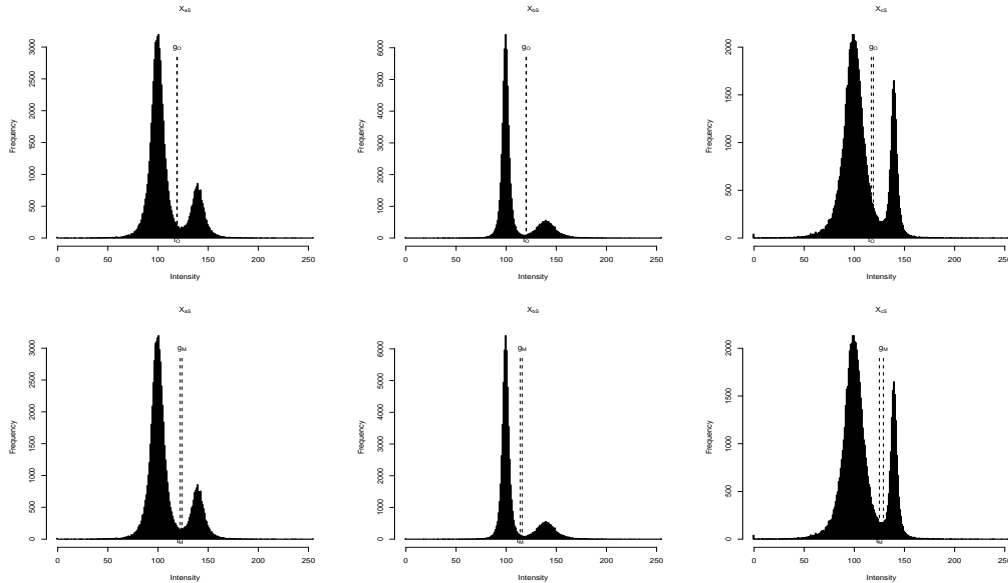


Figure 1: Histograms for  $\mathcal{X}_{aS}$ ,  $\mathcal{X}_{bS}$  and  $\mathcal{X}_{cS}$ , each with two symmetrically distributed classes. Upper row: thresholds  $t_O^*$  and  $g_O^*$ , which were obtained by Otsu's method and its median-based extension, are indicated by dashed lines. Lower row:  $t_M^*$  and  $g_M^*$ , which were obtained by the MET method and its median-based extension, are indicated by dashed lines.

242 The results obtained from applying the four methods in the five scenarios  
 243 ( $S$ ,  $LR$ ,  $LL$ ,  $RR$  and  $RL$ ) of skewness are shown in Figures 1-5, from which  
 244 we can observe the following patterns.

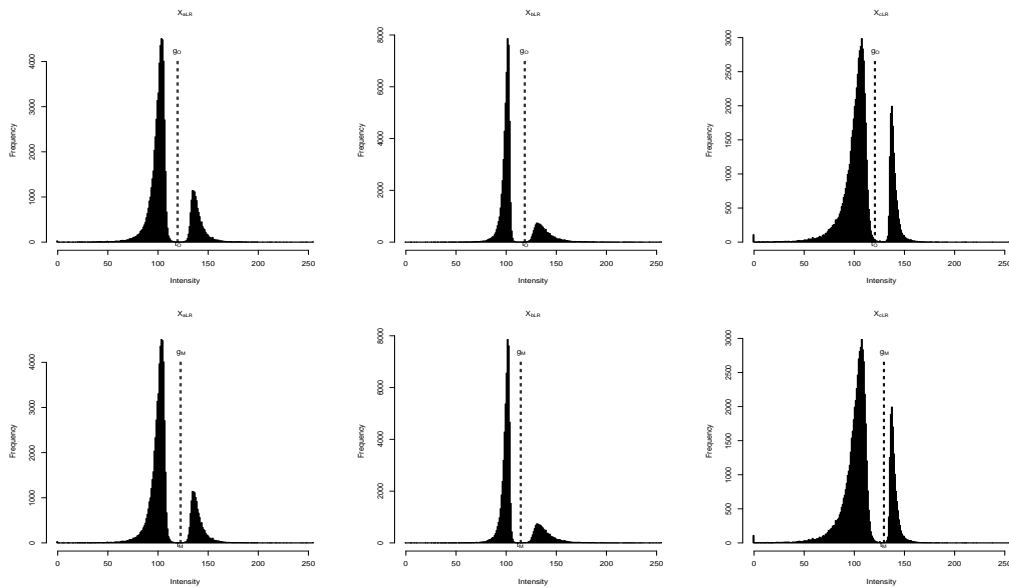


Figure 2: Histograms for  $\mathcal{X}_{aLR}$ ,  $\mathcal{X}_{bLR}$  and  $\mathcal{X}_{cLR}$ , each with two class-conditional distributions skewing away from each other. Caption is as for Figure 1.

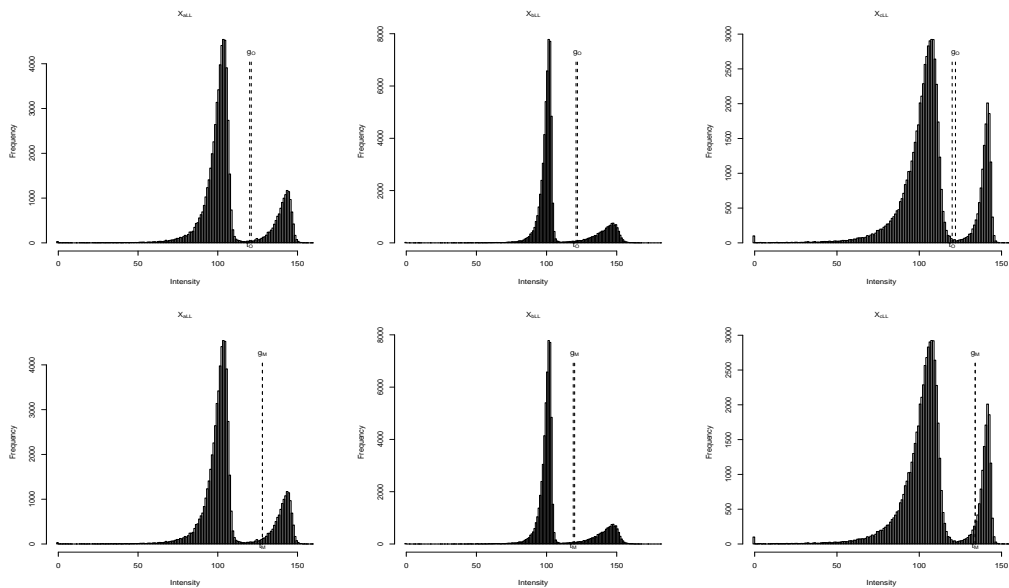


Figure 3: Histograms for  $\mathcal{X}_{aLL}$ ,  $\mathcal{X}_{bLL}$  and  $\mathcal{X}_{cLL}$ , each with the distribution of the minority class skewing towards that of the majority class. Caption is as for Figure 1.

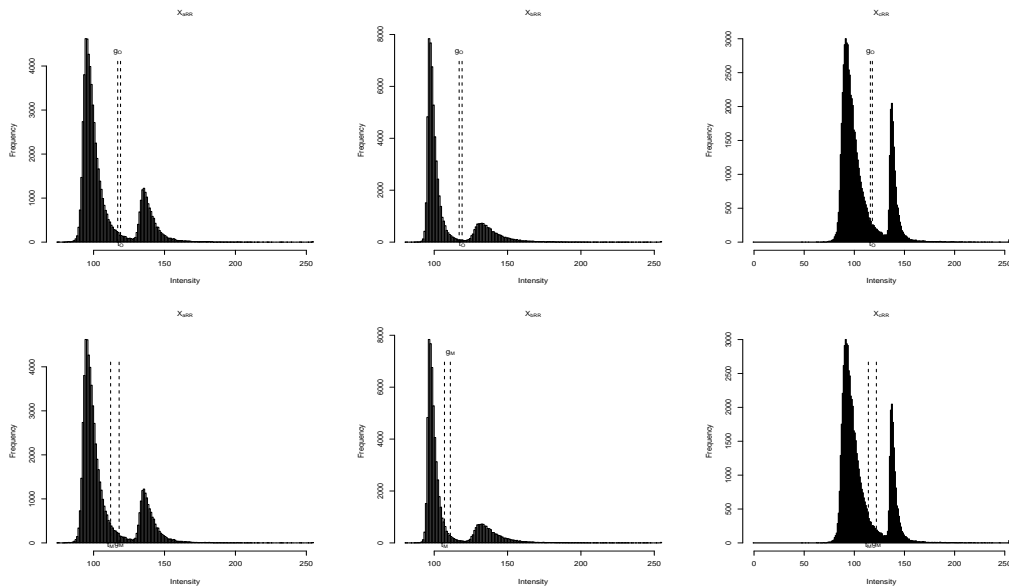


Figure 4: Histograms for  $\mathcal{X}_{aRR}$ ,  $\mathcal{X}_{bRR}$  and  $\mathcal{X}_{cRR}$ , each with the distribution of the majority class skewing towards that of the minority class. Caption is as for Figure 1.

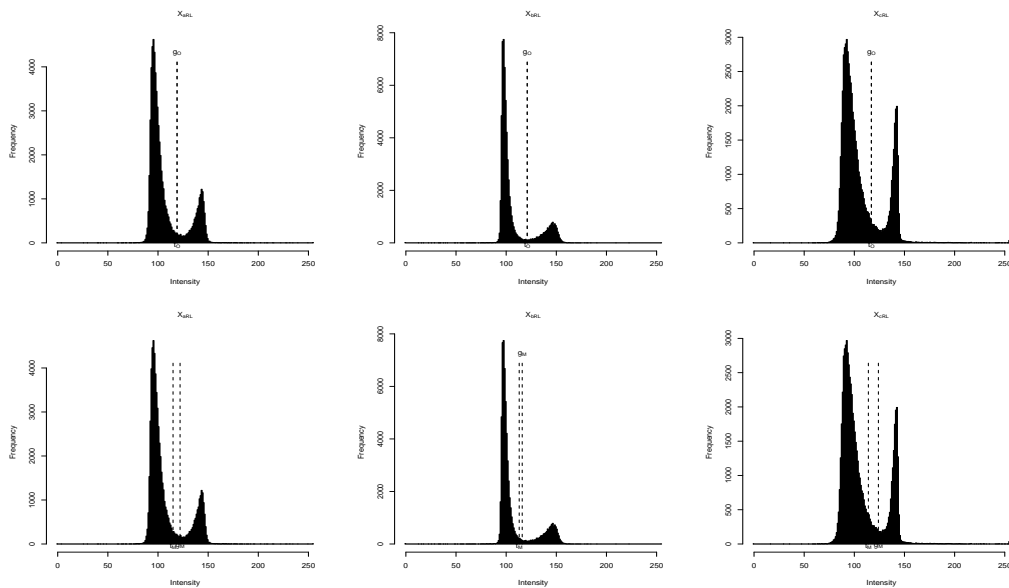


Figure 5: Histograms for  $\mathcal{X}_{aRL}$ ,  $\mathcal{X}_{bRL}$  and  $\mathcal{X}_{cRL}$ , each with two class-conditional distributions skewing towards each other. Caption is as for Figure 1.

245 First, between the two class-conditional distributions there is overlap,  
246 which in our context is mainly due to the skewness and the heavy tails. When  
247 such overlap is not severe, a median-based extension performs similarly to its  
248 original mean-based method. This occurs in either of the first three scenarios:  
249 neither the distribution of the majority class nor that of the minority class is  
250 skew (Figure 1), the two class-conditional distributions skew away from each  
251 other (Figure 2), or the minority class skew towards to the majority class  
252 (Figure 3).

253 Secondly, when such overlap is considerable, the performance of a median-  
254 based extension is comparable to, or noticeably better than, that of its orig-  
255 inal method. Figures 4 and 5 respectively demonstrate respectively for this  
256 case two scenarios: the distribution of the majority class skew towards that  
257 of the minority class, or both class-conditional distributions skew towards  
258 each other.

259 Thirdly, the previous two patterns are valid in our experiments for all  
260 three relationships ( $\mathcal{X}_a$ ,  $\mathcal{X}_b$  and  $\mathcal{X}_c$ ) between the two within-classes variances  
261 for two classes with distinctly unequal sizes.

#### 262 4.2. Real images

263 In addition to the experiments with simulated data sets, we shall illustrate  
264 and compare the performance of Otsu’s method (with  $t_O^*$  or  $\mathbf{t}_O^*$ ), the MET  
265 method (with  $t_M^*$  or  $\mathbf{t}_M^*$ ) and their median-based extensions to binarisation  
266 (with  $g_O^*$  or  $g_M^*$ ) and multi-level thresholding (with  $\mathbf{g}_O^*$  or  $\mathbf{g}_M^*$ ) of real images.  
267 Hence, in this section we present the results obtained from applying the four  
268 methods to six real, test images, which are shown in Figure 6. Of the six  
269 images, three, namely ‘Bonemarr’, ‘Lymp’ and ‘NDT-image20’, were chosen  
270 for binarisation, and the other three, namely ‘Lake’, ‘Girlface’ and ‘NDT-  
271 image18’, were used for tri-level thresholding.

272 These test images are publicly available: ‘Lymp’ is with the software  
273 ImageJ ([rsb.info.nih.gov/ij](http://rsb.info.nih.gov/ij)); ‘Bonemarr’ was with the MATLAB Image Pro-  
274 cessing Toolbox; ‘Lake’ and ‘Girlface’ are two benchmark test images; and  
275 ‘NDT-image20’ and ‘NDT-image18’ are two nondestructive-testing (NDT)  
276 images employed in a recent comprehensive survey of image-thresholding  
277 methods (Sezgin and Sankur, 2004). These images were also presented by  
278 some image binarisation and multi-level thresholding literature including Sa-  
279 hoo and Arora (2004), Sahoo and Arora (2006), Guo and Pandit (1998),  
280 Hammouche et al. (2010) and our recent pieces of work (Xue and Tittering-  
281 ton, 2010a,c).

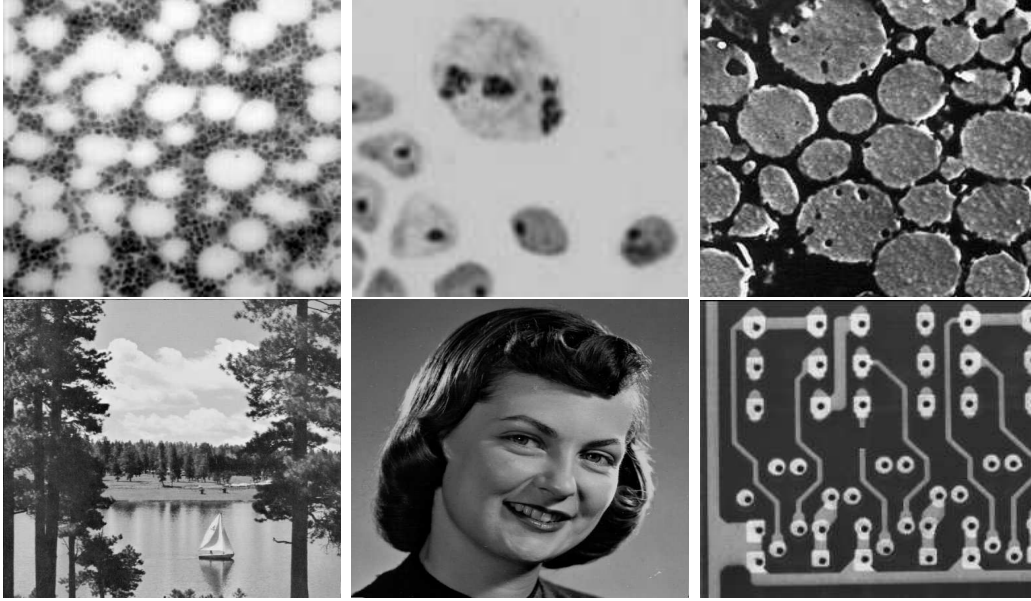


Figure 6: Original real images. Upper row, from left to right: ‘Bonemarr’, ‘Lymp’ and ‘NDT-image20’, for binarisation. Lower row, from left to right: ‘Lake’, ‘Girlface’ and ‘NDT-image18’, for tri-level thresholding.

282 4.2.1. Binarisation

	Bonemarr	Lymp	NDT-image20
$t_O^*$	153	151	79
$g_O^*$	157	166	69
$t_M^*$	175	193	38
$g_M^*$	172	191	39

Table 2: Thresholds  $t_O^*$ ,  $g_O^*$ ,  $t_M^*$  and  $g_M^*$  selected by Otsu’s method and its median-based extension, and the MET method and its median-based extension, respectively, for the binarisation of three real images.

283 The values of thresholds  $t_O^*$ ,  $g_O^*$ ,  $t_M^*$  and  $g_M^*$  for the binarisation of ‘Bone-  
 284 marr’, ‘Lymp’ and ‘NDT-image20’ are listed in Table 2, with the correspond-  
 285 ing binarisation results shown in Figures 7-9.

286 In ‘Bonemarr’, the two classes, for the dark material ( $C_1$ ) and the large  
 287 white cells ( $C_2$ ) respectively, consist of similar numbers of pixels, but their  
 288 intensities are of distinct spreads, as shown in Figure 6 and the histogram in  
 289 Figure 7. In this case, although  $C_1$  is slightly skew and the overlap between

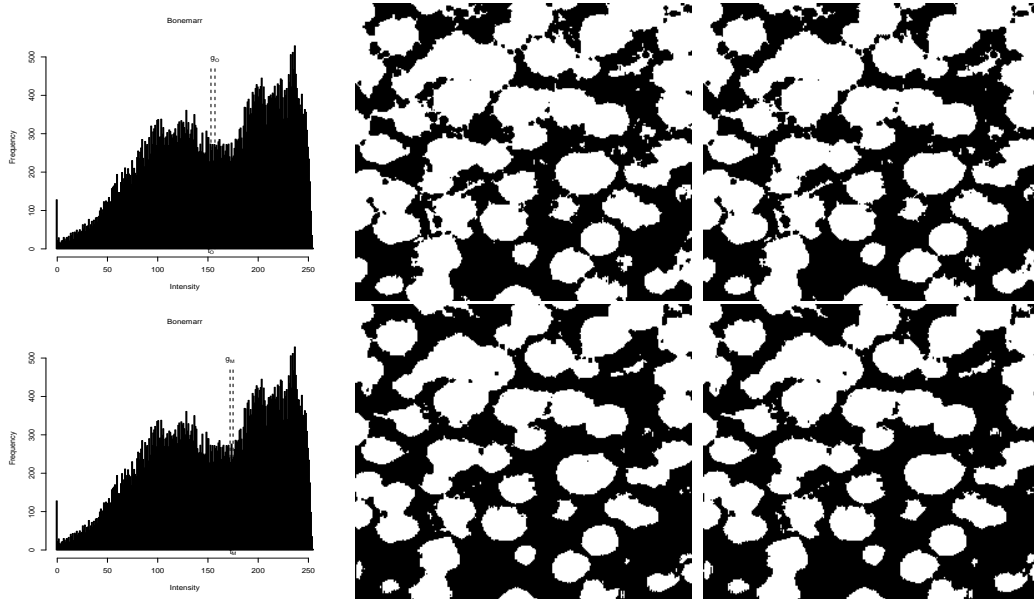


Figure 7: For ‘Bonemarr’. Upper row, from left to right: histogram with thresholds  $t_O^*$  (selected by Otsu’s method) and  $g_O^*$  (by its median-based extension), both indicated by dashed lines; binarised images based on  $t_O^*$  and  $g_O^*$ , respectively. Lower row, from left to right: histogram with thresholds  $t_M^*$  (selected by the MET method) and  $g_M^*$  (by its median-based extension), both indicated by dashed lines; binarised images by using  $t_M^*$  and  $g_M^*$ , respectively.

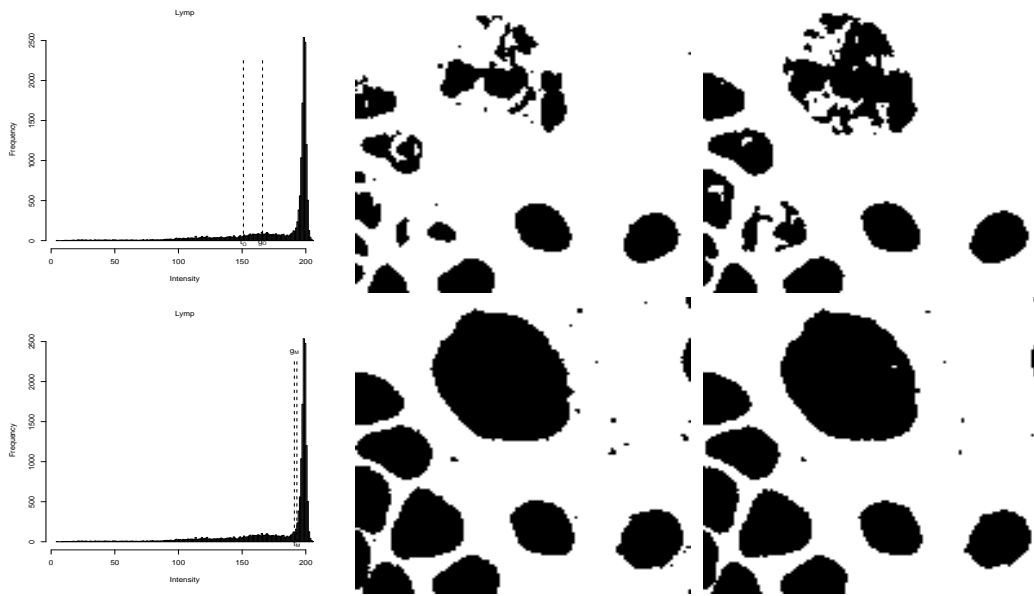


Figure 8: For 'Lymp'. Caption is as for Figure 7.

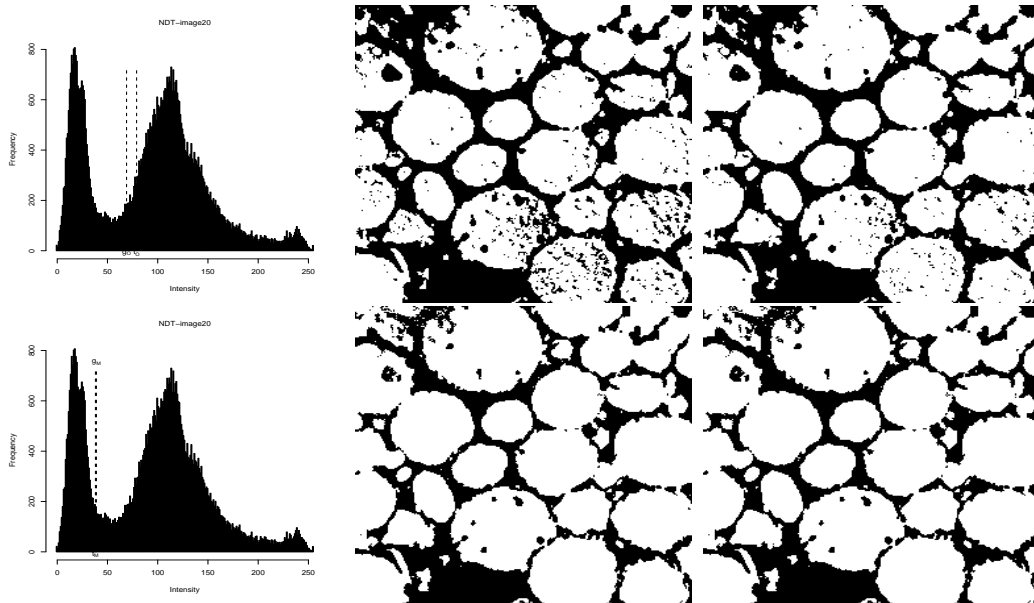


Figure 9: For 'NDT-image20'. Caption is as for Figure 7.

290 the two classes is heavy, each median-based extension performs comparably  
 291 to its original method.

292 As shown in Figures 8 and 9, in both ‘Lymp’ and ‘NDT-image20’ the two  
 293 classes are of distinct sizes and intensity variances. In ‘Lymp’, the minority  
 294 class ( $C_1$  which is for the dark and approximately elliptic cells) has more  
 295 skewed and variable intensities; in ‘NDT-image20’, the majority class ( $C_2$   
 296 which is for the particles) has more skewed and dispersed intensities. In spite  
 297 of the difference, in both cases, the class overlap is moderate; the performance  
 298 of the median-based extension of Otsu’s method is slightly superior to that  
 299 of the original Otsu method; and the performances of the MET method and  
 300 its extension are rather similar.

301 *4.2.2. Tri-level thresholding*

	Lake	Girlface	NDT-image18
$\mathbf{t}_O^*$	(84, 153)	(50, 115)	(94, 170)
$\mathbf{g}_O^*$	(76, 143)	(46, 113)	(94, 172)
$\mathbf{t}_M^*$	(87, 133)	(7, 22)	(32, 63)
$\mathbf{g}_M^*$	(129, 215)	(28, 118)	(21, 69)

Table 3: Thresholds  $\mathbf{t}_O^*$ ,  $\mathbf{g}_O^*$ ,  $\mathbf{t}_M^*$  and  $\mathbf{g}_M^*$  selected by Otsu’s method and its median-based extension, and the MET method and its median-based extension, respectively, for the tri-level thresholding of three real images.

302 The values of thresholds  $\mathbf{t}_O^*$ ,  $\mathbf{g}_O^*$ ,  $\mathbf{t}_M^*$  and  $\mathbf{g}_M^*$  for the tri-level thresholding  
 303 of ‘Lake’, ‘Girlface’ and ‘NDT-image18’ are listed in Table 3.

304 As shown in Figures 10-12, there are three clearly-noticeable modes, and  
 305 one or more classes having skew distributions of intensity, in the histograms  
 306 for ‘Lake’, ‘Girlface’ and ‘NDT-image18’.

307 For ‘Lake’, only the median-based extension of the MET method success-  
 308 fully identified the three classes represented by the three modes of intensity.  
 309 In the case of ‘Girlface’, the two extensions and Otsu’s method succeeded  
 310 in distinguishing the three modes, while the MET method failed. As far  
 311 as ‘NDT-image18’ is concerned, Otsu’s method and its extension work well  
 312 for tri-level thresholding, while neither the MET method, nor its extension,  
 313 separated the two bright classes (i.e. the white rings and the grey lines).

314 In summary, for the binarisation and tri-level thresholding of our test  
 315 images, a median-based extension performs better than, or comparably to,

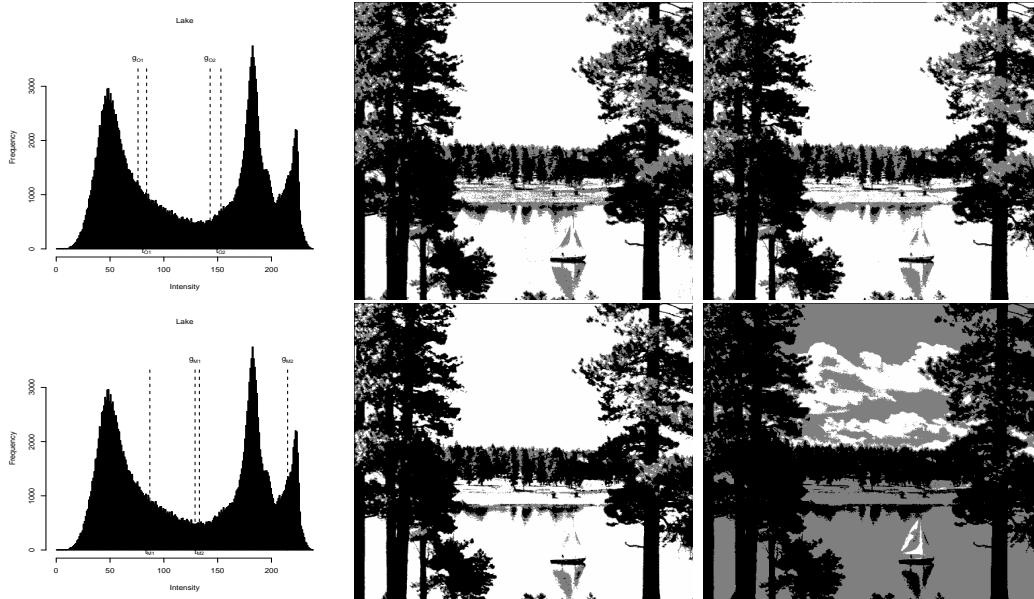


Figure 10: For ‘Lake’. Upper row, from left to right: histogram with thresholds  $\mathbf{t}_O^*$  (selected by Otsu’s method) and  $\mathbf{g}_O^*$  (by its median-based extension), both indicated by dashed lines; tri-level images based on  $\mathbf{t}_O^*$  and  $\mathbf{g}_O^*$ , respectively. Lower row, from left to right: histogram with thresholds  $\mathbf{t}_M^*$  (selected by the MET method) and  $\mathbf{g}_M^*$  (by its median-based extension), both indicated by dashed lines; tri-level images by using  $\mathbf{t}_M^*$  and  $\mathbf{g}_M^*$ , respectively.



Figure 11: For 'Girlface'. Caption is as for Figure 10.

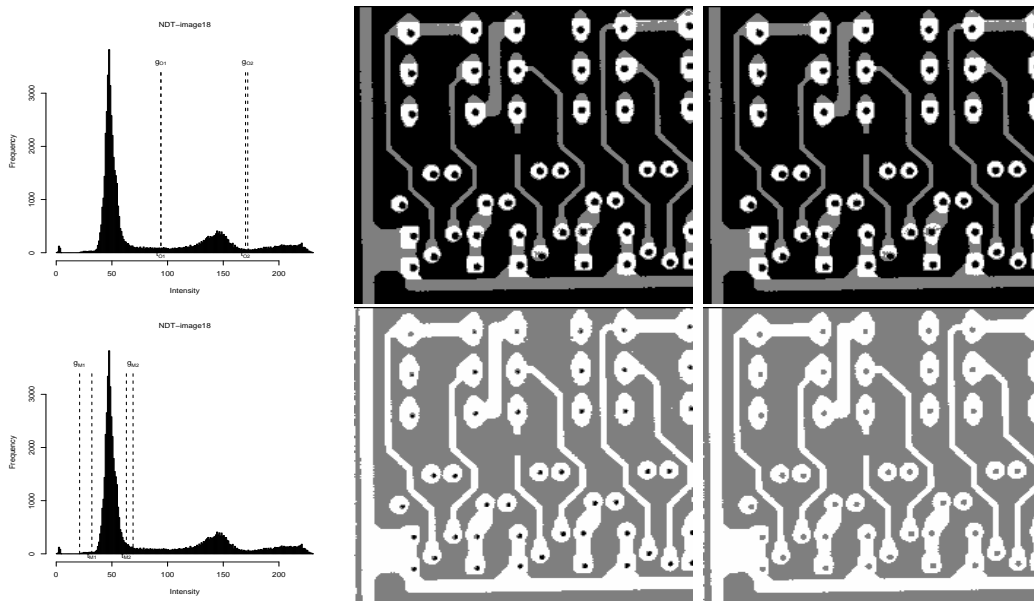


Figure 12: For 'NDT-image18'. Caption is as for Figure 10.

316 its original version, indicating a more robust performance to the presence of  
317 skew class-conditional distributions.

## 318 5. Conclusion

319 In this paper, we proposed two median-based approaches to image thresh-  
320 olding. One approach is an extension of Otsu's method, and the other is an  
321 extension of Kittler and Illingworth's MET method. Both extensions were  
322 aimed to improve the robustness of their original methods to the presence  
323 of skew or heavy-tailed class-conditional distributions, while preserving the  
324 methodological simplicity and computational efficiency of the originals. Our  
325 experiments on some simulated data from skew  $t$ -distributions and some real  
326 images demonstrated that the two extensions could accomplish more robust  
327 performance than that of the originals. In addition, we provided theoret-  
328 ical interpretation of the new approaches, based on the mixture of Laplace  
329 distributions.

## 330 Acknowledgments

331 The NDT images are provided through the courtesy of Dr. Mehmet  
332 Sezgin. This work was partly supported by funding to J.-H.X. from the  
333 Internal Visiting Programme, under the EU-funded PASCAL2 Network of  
334 Excellence.

## 335 References

- 336 Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of  
337 symmetry with emphasis on a multivariate skew  $t$ -distribution. *Journal of*  
338 *the Royal Statistical Society: Series B* 65 (2), 367–389.
- 339 Bazi, Y., Bruzzone, L., Melgani, F., 2007. Image thresholding based on the  
340 EM algorithm and the generalized Gaussian distribution. *Pattern Recogn-*  
341 *ition* 40 (2), 619–634.
- 342 Fan, S.-K. S., Lin, Y., Wu, C.-C., 2008. Image thresholding using a novel  
343 estimation method in generalized Gaussian distribution mixture modeling.  
344 *Neurocomputing* 72 (1-3), 500–512.
- 345 Glasbey, C. A., 1993. An analysis of histogram-based thresholding algo-  
346 rithms. *CVGIP: Graphical Models and Image Processing* 55 (6), 532–537.

- 347 Guo, R., Pandit, S. M., 1998. Automatic threshold selection based on his-  
348 togram modes and a discriminant criterion. *Machine Vision and Applica-*  
349 *tions* 10 (5-6), 331–338.
- 350 Hammouche, K., Diaf, M., Siarry, P., 2010. A comparative study of various  
351 meta-heuristic techniques applied to the multilevel thresholding problem.  
352 *Engineering Applications of Artificial Intelligence* 23 (5), 676–688.
- 353 Kittler, J., Illingworth, J., 1986. Minimum error thresholding. *Pattern Recog-*  
354 *nition* 19 (1), 41–47.
- 355 Kurita, T., Otsu, N., Abdelmalek, N., 1992. Maximum likelihood thresh-  
356 olding based on population mixture models. *Pattern Recognition* 25 (10),  
357 1231–1240.
- 358 Moser, G., Serpico, S. B., 2006. Generalized minimum-error thresholding  
359 for unsupervised change detection from SAR amplitude imagery. *IEEE*  
360 *Transactions on Geoscience and Remote Sensing* 44 (10), 2972–2982.
- 361 Otsu, N., 1979. A threshold selection method from gray-level histograms.  
362 *IEEE Transactions on Systems, Man, and Cybernetics* SMC-9, 62–66.
- 363 Pal, N. R., Bhandari, D., 1993. Image thresholding: some new techniques.  
364 *Signal Processing* 33 (2), 139–158.
- 365 Sahoo, P. K., Arora, G., 2004. A thresholding method based on two-  
366 dimensional Renyi’s entropy. *Pattern Recognition* 37 (6), 1149–1161.
- 367 Sahoo, P. K., Arora, G., 2006. Image thresholding using two-dimensional  
368 Tsallis-Havrda-Charvát entropy. *Pattern Recognition Letters* 27 (6), 520–  
369 528.
- 370 Sahoo, P. K., Soltani, S., Wong, A. K. C., Chen, Y. C., 1988. A survey of  
371 thresholding techniques. *Computer Vision, Graphics, and Image Process-*  
372 *ing* 41 (2), 233–260.
- 373 Sezgin, M., Sankur, B., 2004. Survey over image thresholding techniques and  
374 quantitative performance evaluation. *Journal of Electronic Imaging* 13 (1),  
375 146–165.

- 376 Trier, Ø. D., Jain, A. K., 1995. Goal-directed evaluation of binarization  
377 methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*  
378 17 (12), 1191–1201.
- 379 Xue, J.-H., Titterington, D. M., 2010a. Image thresholding by using statistics  
380 for determining the number of clusters. manuscript.
- 381 Xue, J.-H., Titterington, D. M., 2010b.  $t$ -tests,  $F$ -tests and Otsu’s methods  
382 for image thresholding. manuscript.
- 383 Xue, J.-H., Titterington, D. M., 2010c. Threshold selection from image his-  
384 tograms with skewed components based on maximum-likelihood estimation  
385 of skew-normal and log-concave distributions. manuscript.
- 386 Xue, J.-H., Zhang, Y. J., Lin, X. G., 1999. Rayleigh-distribution based min-  
387 imum error thresholding for SAR images. *Journal of Electronics (China)*  
388 16 (4), 336–342.
- 389 Yan, H., 1996. Unified formulation of a class of image thresholding techniques.  
390 *Pattern Recognition* 29 (12), 2025–2032.