

Image Thresholding by Using Statistics for Determining the Number of Clusters

Jing-Hao Xue^{a,*}, D. Michael Titterington^b

^a*Department of Statistical Science, University College London, London WC1E 6BT, UK*

^b*School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8QQ, UK*

Abstract

This paper aims to demonstrate that some statistics originally developed in cluster analysis for determining the number of clusters can be instead appropriately used in image thresholding to select optimal thresholds. Three highly-cited statistics, namely the Calinski-Harabasz index, the Gap statistic and the Silhouette widths, are chosen and applied to real images and simulated data, for illustrative purposes. The relationship between these statistics and a commonly-used image-thresholding method, Otsu's method, is also briefly discussed.

Keywords: Calinski-Harabasz index, Cluster analysis, Cluster validity, Gap statistic, Image segmentation, Image thresholding, Otsu's method, Silhouette widths, Unsupervised learning.

1. Introduction

2 The task of image segmentation is to partition an image into a number
3 of mutually-exclusive segments with each segment belonging to one of K
4 predetermined groups. In its simplest and most common case, the image is
5 partitioned into segments of two groups, the background and the foreground,
6 i.e. $K = 2$.

7 One of the techniques for image segmentation is image thresholding,
8 which partitions the range of intensities (or so-called gray levels) of image

*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 pixels into K mutually-exclusive intervals, by using $K - 1$ thresholds. This
10 technique is valid if the groups are distinct from each other in their inten-
11 sities. Such a simple technique is widely used in practice and has a long
12 history of research effort. This effort has resulted in a considerable number
13 of thresholding methods, of which nice surveys and comparative investiga-
14 tions can be found in Sahoo et al. (1988), Glasbey (1993), Trier and Jain
15 (1995) and Sezgin and Sankur (2004), among others.

16 Image segmentation, usually with no group label being given to any pixel
17 a priori, is an unsupervised learning task, and thus it can be accomplished
18 through cluster analysis (Jain et al., 1999).

19 In cluster analysis, one of the fundamental and the most-difficult problems
20 is to determine K , the number of groups (or clusters) (Xu and Wunsch, 2005;
21 Jain, 2010). In order to select an optimal K , tens if not hundred of statistics
22 have been developed to evaluate and compare clustering results over possible
23 values of K ; see comprehensive surveys and comparative studies by Xu and
24 Wunsch (2005), Milligan and Cooper (1985), Chiang and Mirkin (2010) and
25 Jain (2010), among others.

26 The aim, and the main contribution, of this paper is to demonstrate
27 that some statistics, which were originally developed in cluster analysis for
28 determining K , can instead be used in image thresholding to select an optimal
29 threshold t , or an optimal set of thresholds \mathbf{t} when K is larger than two.

30 For illustrative purposes, three highly-cited statistics, namely the Calinski-
31 Harabasz index (Milligan and Cooper, 1985), the Gap statistic (Tibshirani
32 et al., 2001) and the Silhouette widths (Rousseeuw, 1987), are chosen and ap-
33 plied to real images and simulated data. The relationship between the statis-
34 tics and a commonly-used image-thresholding method, Otsu’s method, is also
35 briefly discussed. Some investigation into using many other K -determination
36 statistics for image thresholding, or more general image segmentation, may
37 be encouraged by the results presented in this paper.

38 2. Methods

39 2.1. Image thresholding

40 Consider an image \mathcal{X} of N pixels with each pixel represented by its inten-
41 sity $x_i, i = 1, \dots, N$. A strictly-monotonically-increasing sequence of $K - 1$
42 thresholds, $\mathbf{t} = (t_1, \dots, t_{K-1})$, partitions the range $[0, T)$ of intensities in \mathcal{X} ,
43 and thus \mathcal{X} itself, into K mutually-exclusive groups $\mathcal{C}_1(\mathbf{t}), \dots, \mathcal{C}_K(\mathbf{t})$, where
44 $\mathcal{C}_k(\mathbf{t}) = \{i : t_{k-1} \leq x_i < t_k, 1 \leq i \leq N\}$ with $k = 1, \dots, K$, $t_0 = 0$ and

45 $t_K = T$. Therefore, the task of image thresholding is to select an optimal \mathbf{t}
 46 (denoted by \mathbf{t}^* hereafter), for an often-predetermined K .

47 As mentioned in section 1, we shall use the Calinski-Harabasz index, the
 48 Gap statistic and the Silhouette widths, respectively, to select \mathbf{t}^* . These
 49 statistics were originally developed in cluster analysis for determining K , the
 50 number of groups.

51 2.2. The Calinski-Harabasz index

52 Using four hierarchical-clustering methods based on the single link, com-
 53 plete link, group average and Ward's minimum variance, respectively, Milli-
 54 gan and Cooper (1985) compare thirty indices or procedures for determin-
 55 ing K . From this comprehensive study with simulated data, the Calinski-
 56 Harabasz index (CHI) was observed as the best statistic for determining K .

57 The CHI can be calculated as

$$\text{CHI}(K) = \frac{SS_B(K)/(K-1)}{SS_W(K)/(N-K)}, \quad (1)$$

58 where $SS_W(K)$ and $SS_B(K)$ are the within-group and between-group sums
 59 of squares, respectively, after the data have been divided into K groups. An
 60 optimal K (denoted by K^* hereafter), is obtained through maximisation of
 61 $\text{CHI}(K)$ over possible values of K .

62 It is apparent that the CHI is equal to the F statistic that is used in the
 63 one-way analysis of variance (ANOVA) to compare K groups of data. In
 64 other words, K^* is the K that provides the largest F statistic.

65 In image thresholding, the two sums of squares SS_W and SS_B , and thus
 66 the CHI, are functions of not only K but also thresholds \mathbf{t} ; hence, we rewrite
 67 them as $SS_W(K, \mathbf{t})$, $SS_B(K, \mathbf{t})$ and $\text{CHI}(K, \mathbf{t})$ for explicitness.

68 As the CHI is a function of both K and \mathbf{t} , it can be used to select K^*
 69 and \mathbf{t}^* simultaneously:

$$(K^*, \mathbf{t}^*) = \underset{K, \mathbf{t}}{\text{argmax}} \text{CHI}(K, \mathbf{t}) = \underset{K, \mathbf{t}}{\text{argmax}} \frac{SS_B(K, \mathbf{t})/(K-1)}{SS_W(K, \mathbf{t})/(N-K)}. \quad (2)$$

70 For an image-thresholding task with a fixed K (e.g. in the common case
 71 of image binarisation for which $K = 2$), it follows that the CHI can be used
 72 to select \mathbf{t}^* :

$$\mathbf{t}_C^* = \underset{\mathbf{t}}{\text{argmax}} \text{CHI}(\mathbf{t}) = \underset{\mathbf{t}}{\text{argmax}} \frac{SS_B(\mathbf{t})/(K-1)}{SS_W(\mathbf{t})/(N-K)}. \quad (3)$$

73 This criterion, as both N and K are fixed, is equivalent to one of the most
74 widely-used thresholding methods, Otsu’s method (Otsu, 1979), which se-
75 lects \mathbf{t}^* that provides the maximum of the ratio between $SS_B(\mathbf{t})$ and $SS_W(\mathbf{t})$.
76 Otsu’s method is used by some commercial and free software, such as MAT-
77 LAB (The MathWorks, Inc.) and GIMP (www.gimp.org), as the method for
78 automatic image thresholding.

79 As the sum, $SS_W(\mathbf{t}) + SS_B(\mathbf{t})$, is a constant invariant to \mathbf{t} , the criterion
80 in equation (3) is also equivalent to maximisation of $SS_B(\mathbf{t})$ or minimisation
81 of $SS_W(\mathbf{t})$ (Otsu, 1979; Kurita et al., 1992):

$$\mathbf{t}_C^* = \underset{\mathbf{t}}{\operatorname{argmax}} \operatorname{CHI}(\mathbf{t}) = \underset{\mathbf{t}}{\operatorname{argmax}} SS_B(\mathbf{t}) = \underset{\mathbf{t}}{\operatorname{argmin}} SS_W(\mathbf{t}) . \quad (4)$$

82 In this paper, mainly for reasons of consistency and comparison with two
83 other methods to be described shortly, we adopt

$$\mathbf{t}_C^* = \underset{\mathbf{t}}{\operatorname{argmax}} \operatorname{CHI}(\mathbf{t}) = \underset{\mathbf{t}}{\operatorname{argmax}} \{-SS_W(\mathbf{t})/N\} . \quad (5)$$

84 2.3. The Gap statistic

85 The Gap statistic, developed by Tibshirani et al. (2001), is another highly-
86 cited statistic for determining K in cluster analysis.

87 The Gap statistic is a measure of the distance between the within-group
88 sum of squares of the observed data and that of the data sampled from an
89 appropriate null reference distribution. Hence, the K corresponding to the
90 largest Gap statistic can be selected as K^* . The Gap statistic is defined as

$$\operatorname{Gap}(K) = \frac{1}{B} \sum_{b=1}^B \log\{SS_W^b(K)\} - \log\{SS_W(K)\} , \quad (6)$$

91 where $SS_W^b(K)$, $b = 1, \dots, B$, is the within-group sum of squares for the b -th
92 data set sampled from the reference distribution. In Tibshirani et al. (2001),
93 a uniform distribution over the range of the observed data has been used as
94 the reference distribution.

95 Instead of using the largest Gap statistic, Tibshirani et al. (2001) select
96 as K^* the smallest value of K such that $\operatorname{Gap}(K) \leq \operatorname{Gap}(K+1) - s_{K+1}$, where
97 s_{K+1} is an estimator of the standard deviation of $\log(SS_W^b(K+1))$. This
98 selection is based on the consideration of sampling variation associated with
99 the B data sets and on the so-called ‘one-standard-error’ rule for selecting
100 a parsimonious model; empirically it works well (Tibshirani et al., 2001) for

101 determining K . However, it involves performing cluster analysis for both K
 102 and $K + 1$ groups. In image thresholding for K groups, it is desired not
 103 to investigate a more complex case of $K + 1$ groups. Hence, we adopt the
 104 criterion of using the largest Gap statistic for determining \mathbf{t} .

105 As with the CHI, the $B + 1$ sums of squares SS_W^b and SS_W , and thus the
 106 Gap statistic, are functions of not only K but also \mathbf{t} . Therefore, we can use
 107 $\text{Gap}(K, \mathbf{t})$ to select K^* and \mathbf{t}^* simultaneously as

$$\begin{aligned} (K^*, \mathbf{t}^*) &= \underset{K, \mathbf{t}}{\operatorname{argmax}} \text{Gap}(K, \mathbf{t}) \\ &= \underset{K, \mathbf{t}}{\operatorname{argmax}} \left[\frac{1}{B} \sum_{b=1}^B \log\{SS_W^b(K, \mathbf{t})\} - \log\{SS_W(K, \mathbf{t})\} \right], \quad (7) \end{aligned}$$

108 and, for image thresholding with a fixed K , select \mathbf{t}^* as

$$\begin{aligned} \mathbf{t}_G^* &= \underset{\mathbf{t}}{\operatorname{argmax}} \text{Gap}(\mathbf{t}) \\ &= \underset{\mathbf{t}}{\operatorname{argmax}} \left[\frac{1}{B} \sum_{b=1}^B \log\{SS_W^b(\mathbf{t})\} - \log\{SS_W(\mathbf{t})\} \right]. \quad (8) \end{aligned}$$

109 To use equation (8), we need to sample a uniform reference distribution
 110 such that B data sets can be obtained to calculate $\log\{SS_W^b(\mathbf{t})\}$. In order
 111 to derive a simplified, non-sampling way of selecting \mathbf{t}^* , we consider the fact
 112 that image thresholding is usually a univariate technique; that is, we use a
 113 single feature, the pixel intensity. The average $\log SS_W^b(\mathbf{t})$ is an estimator
 114 of the logarithm of the within-interval sum of squares for a partition of the
 115 range $[0, T)$ into K intervals, when for each intensity there are N/T pixels.
 116 Hence, based on the calculation of the sum of squares for a discrete uniform
 117 distribution, a non-sampling approximation to equation (8) can be derived
 118 as

$$\begin{aligned} \mathbf{t}_G^* &= \underset{\mathbf{t}}{\operatorname{argmax}} \text{Gap}(\mathbf{t}) \\ &= \underset{\mathbf{t}}{\operatorname{argmax}} \left[\log \left\{ \frac{N}{T} \sum_{k=1}^K \frac{(t_k - t_{k-1})^3 - (t_k - t_{k-1})}{12} \right\} \right. \\ &\quad \left. - \log\{SS_W(\mathbf{t})\} \right], \quad (9) \end{aligned}$$

119 where, as defined before, $\mathbf{t} = (t_1, \dots, t_{K-1})$, $t_0 = 0$ and $t_K = T$.

120 Tibshirani et al. (2001) mention that $\log\{SS_W\}$ has an interpretation as
 121 a log-likelihood based on a special Gaussian mixture. Indeed, maximisation
 122 of $-\log\{SS_W(\mathbf{t})\}$ for image thresholding is equivalent to maximisation of
 123 log-likelihood based on a Gaussian mixture with equal variances and equal
 124 proportions across the component (Gaussian) distributions, for any candidate
 125 threshold \mathbf{t} , and it is equivalent to Otsu’s method (Kurita et al., 1992).

126 However, given that there is also a term associated with $SS_W^b(\mathbf{t})$ along
 127 with $-\log\{SS_W(\mathbf{t})\}$, equation (9) may provide a \mathbf{t}^* different from that ob-
 128 tained by Otsu’s method.

129 2.4. The Silhouette widths

130 The last highly-cited statistic, originally developed for determining K ,
 131 that we shall use in this paper for determining \mathbf{t} , is called the Silhouette
 132 widths (Rousseeuw, 1987).

133 In the context of image segmentation, the Silhouette width for the i -th
 134 pixel (represented by its intensity x_i) is proposed to measure the difference
 135 between its dissimilarity from other pixels in the same group and its dissim-
 136 ilarity from the pixels in the ‘nearest-neighbour’ group. The dissimilarity
 137 (or distance) between pixels is usually quantified by the difference in their
 138 intensities.

139 For x_i , its Silhouette width is defined as

$$s(x_i, K) = \frac{b(x_i, K) - a(x_i, K)}{\max\{a(x_i, K), b(x_i, K)\}}, \quad (10)$$

140 where $a(x_i, K)$ is the average distance from x_i to all other pixels in the same
 141 group, and $b(x_i, K)$ is the minimum of the average distances from x_i to all
 142 pixels in each of the other groups. In other words, $b(x_i, K)$ is the average
 143 distance from x_i to all other pixels in the ‘nearest-neighbour’ group. The
 144 value of $s(x_i, K)$ lies in the range $[-1, 1]$, and a larger value implies a better
 145 segmentation result.

146 Therefore, the average Silhouette width over pixels within a group can
 147 be used to assess the group’s compactness and its separateness from other
 148 groups. The overall average Silhouette width over all pixels in an image can
 149 be used to select K^* by maximisation of it over K .

150 In image thresholding, $a(x_i, K)$, $b(x_i, K)$ and thus $s(x_i, K)$ all vary with

151 \mathbf{t} . As with the CHI and the Gap statistic, we have

$$\begin{aligned}
(K^*, \mathbf{t}^*) &= \operatorname{argmax}_{K, \mathbf{t}} \text{SI}(K, \mathbf{t}) \\
&= \operatorname{argmax}_{K, \mathbf{t}} \frac{1}{N} \sum_{i=1}^N s(x_i, K, \mathbf{t}) \\
&= \operatorname{argmax}_{K, \mathbf{t}} \frac{1}{N} \sum_{i=1}^N \frac{b(x_i, K, \mathbf{t}) - a(x_i, K, \mathbf{t})}{\max\{a(x_i, K, \mathbf{t}), b(x_i, K, \mathbf{t})\}}, \quad (11)
\end{aligned}$$

152 and, for a fixed K , an optimal \mathbf{t}^* can be defined as

$$\begin{aligned}
\mathbf{t}_S^* &= \operatorname{argmax}_{\mathbf{t}} \text{SI}(\mathbf{t}) = \operatorname{argmax}_{\mathbf{t}} \frac{1}{N} \sum_{i=1}^N s(x_i, \mathbf{t}) \\
&= \operatorname{argmax}_{\mathbf{t}} \frac{1}{N} \sum_{i=1}^N \frac{b(x_i, \mathbf{t}) - a(x_i, \mathbf{t})}{\max\{a(x_i, \mathbf{t}), b(x_i, \mathbf{t})\}}. \quad (12)
\end{aligned}$$

153 Although the overall average Silhouette width, $\text{SI}(\mathbf{t})$, and the CHI, $\text{CHI}(\mathbf{t})$,
154 both consider two characteristics, namely compactness and separateness, of
155 the groups, they measure and combine the two characteristics in different
156 ways. The use of $a(x_i, \mathbf{t})$ and $b(x_i, \mathbf{t})$ and the combination of them makes the
157 Silhouette-width-based thresholding method take both characteristics into
158 consideration, as indicated by equation (12). In contrast, the use of the
159 within-group sum of squares $SS_W(\mathbf{t})$ and the between-group sum of squares
160 $SS_B(\mathbf{t})$ makes the CHI-based thresholding method only consider the com-
161 pactness of the groups, as indicated by equation (5). The Gap-statistic-
162 based thresholding method mainly considers the compactness, measured by
163 $\log\{SS_W(\mathbf{t})\}$, as indicated by equation (9).

164 2.4.1. Fast computation of the Silhouette widths for image thresholding

165 In order to calculate the overall average Silhouette width $\text{SI}(t)$, we need to
166 compute $a(x_i, \mathbf{t})$ and $b(x_i, \mathbf{t})$ for each pixel x_i . This makes the computation
167 much more demanding than those for the CHI and the Gap statistic. To
168 store an $N \times N$ distance matrix is also an issue, where $N = 65536$ for an
169 image of 256×256 pixels. Fortunately, for image thresholding, $\text{SI}(t)$ can be
170 calculated fast by only computing $a(x, \mathbf{t})$ and $b(x, \mathbf{t})$ for each intensity x , and
171 it is only necessary to store an $T \times T$ distance matrix at most, where $T = 256$

172 for an 8-bit gray-level image. The reason is that we can rewrite equation (12)
 173 as

$$\begin{aligned} \mathbf{t}_S^* &= \operatorname{argmax}_{\mathbf{t}} \operatorname{SI}(\mathbf{t}) = \operatorname{argmax}_{\mathbf{t}} \frac{1}{N} \sum_{x=0}^{T-1} \{h(x)s(x, \mathbf{t})\} \\ &= \operatorname{argmax}_{\mathbf{t}} \frac{1}{N} \sum_{x=1}^{T-1} \left\{ h(x) \frac{b(x, \mathbf{t}) - a(x, \mathbf{t})}{\max\{a(x, \mathbf{t}), b(x, \mathbf{t})\}} \right\}, \end{aligned} \quad (13)$$

174 where $h(x)$ is the frequency (i.e. the number of pixels) of the intensity x in
 175 the image \mathcal{X} , and, for $x \in \mathcal{C}_k(\mathbf{t})$ (i.e. $t_{k-1} \leq x < t_k$), $a(x, \mathbf{t})$ and $b(x, \mathbf{t})$ can
 176 be computed as follows:

$$a(x, \mathbf{t}) = \frac{1}{N_k(\mathbf{t})} \sum_{g \in \mathcal{C}_k(\mathbf{t})} \{h(g)|x - g|\}, \quad (14)$$

177

$$b(x, \mathbf{t}) = \min_{l \in \{k-1, k+1\}} \frac{1}{N_l(\mathbf{t})} \sum_{g \in \mathcal{C}_l(\mathbf{t})} \{h(g)|x - g|\}, \quad (15)$$

178 in which $N_l(\mathbf{t}) = \sum_{g \in \mathcal{C}_l(\mathbf{t})} h(g)$ and the Euclidean distance is adopted.

179 Only groups $\mathcal{C}_{k-1}(\mathbf{t})$ and $\mathcal{C}_{k+1}(\mathbf{t})$ are considered in equation (15); this is
 180 because, for an intensity x located in $\mathcal{C}_k(\mathbf{t})$, its nearest-neighbour group is
 181 either $\mathcal{C}_{k-1}(\mathbf{t})$ or $\mathcal{C}_{k+1}(\mathbf{t})$ in image thresholding. This can be further sim-
 182 plified for image binarisation when there exist only two groups $\mathcal{C}_1(t_1)$ and
 183 $\mathcal{C}_2(t_1)$, which are the nearest-neighbour groups to each other. In this way,
 184 the calculation for $\operatorname{SI}(t_1)$ is as fast as, if not faster than, those for $\operatorname{CHI}(t_1)$
 185 and $\operatorname{Gap}(t_1)$ in our experiments below.

186 3. Experimental Studies

187 For illustration, in this section we only deal with image binarisation (i.e.
 188 $K = 2$), the most-common case of image thresholding, for simulated data
 189 and real images. To select a single threshold t^* (i.e. t_1^* of \mathbf{t}^*) for these data
 190 sets, we use equations (5), (9) and (13), in which t_C^* , t_G^* and t_S^* correspond
 191 to the thresholds selected by the use of the CHI, the Gap statistic and the
 192 overall average Silhouette width, respectively.

193 *3.1. Simulated histograms*

194 Similarly to Kurita et al. (1992) and Glasbey (1993), we build various data
 195 sets, with each set containing simulated data from a two-component Gaussian
 196 mixture. Hence, the data in different sets correspond to pixel intensities in
 197 different types of virtual image of two groups (i.e. the background and the
 198 foreground). In addition, in a virtual image \mathcal{X} , the pixel intensity of one
 199 group follows a Gaussian $N(\mu_1, \sigma_1^2)$ distribution and that of the other group
 200 follows $N(\mu_2, \sigma_2^2)$. The group proportions are π_1 and π_2 with $\pi_1 + \pi_2 = 1$.

201 We set $T = 256$, $\mu_1 = 100$, $\mu_2 = 150$, and $N = 256 \times 256$ as the size of
 202 \mathcal{X} . Since intensities are integers in the range $[0, T)$, we round, left-truncate
 203 and right-censor the simulated data into $[0, T)$.

204 In this setting, π_1 can be larger than, equal to or smaller than π_2 , and
 205 there are similar relationships between σ_1 and σ_2 ; their combinations lead to
 206 nine types of simulated data and thus histograms constructed from the data.
 207 Since some pairs of types are mirrors of each other, we retain five unique
 208 types with parameter values shown in Table 1, and correspondingly build
 five sets (denoted by $\mathcal{X}_1, \dots, \mathcal{X}_5$ hereafter), of simulated data.

	π_1	π_2	σ_1	σ_2	t_C^*	t_G^*	t_S^*
\mathcal{X}_1	0.50	0.50	10	10	125	125	125
\mathcal{X}_2	0.50	0.50	5	15	127	127	125
\mathcal{X}_3	0.95	0.05	10	10	123	121	138
\mathcal{X}_4	0.95	0.05	5	15	126	125	132
\mathcal{X}_5	0.95	0.05	15	5	107	104	139

Table 1: Parameters of two-component Gaussian mixtures for five simulated data sets with corresponding thresholds t_C^* , t_G^* and t_S^* selected by the use of the CHI, the Gap statistic and the overall average Silhouette width, respectively.

209
 210 For these five data sets, the statistics $\text{CHI}(t)$, $\text{Gap}(t)$ and $\text{SI}(t)$, all of
 211 which have been rescaled to the range $[0, 1]$, are plotted versus t in the left-
 212 hand panels of Figures 1-5. From these plots, we can observe that $\text{CHI}(t)$ is
 213 often, although not necessarily, unimodal with the mode located away from
 214 the ends of the intensity range (i.e. t_0 or t_K), whereas $\text{Gap}(t)$ and $\text{SI}(t)$, in
 215 particular $\text{SI}(t)$, may have their global maxima located at t_0 or t_K .

216 Equation (9) shows that the Gap-statistic-based method may pull the
 217 threshold obtained by the CHI-based method (i.e. Otsu's method) further
 218 away from the middle of the intensity range.

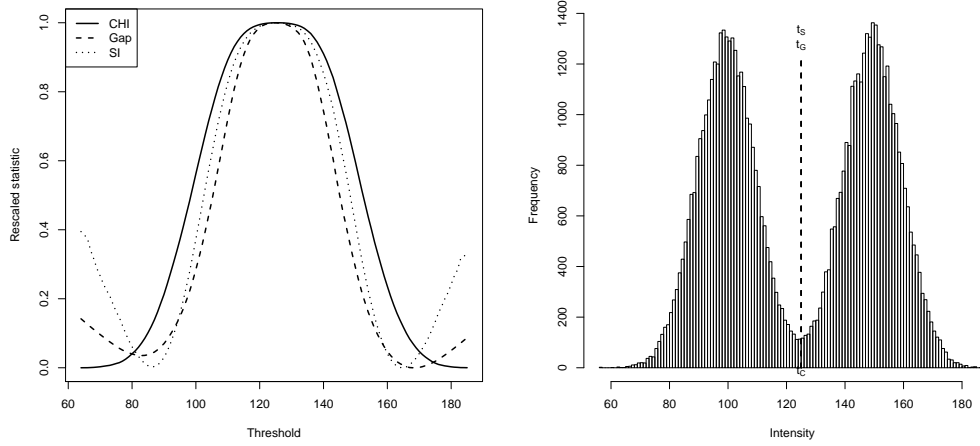


Figure 1: For \mathcal{X}_1 . Left-hand panel: $\text{CHI}(t)$, $\text{Gap}(t)$ and $\text{SI}(t)$, all of which have been rescaled to the range $[0, 1]$; right-hand panel: histogram with thresholds t_C^* , t_G^* and t_S^* indicated by dashed lines.

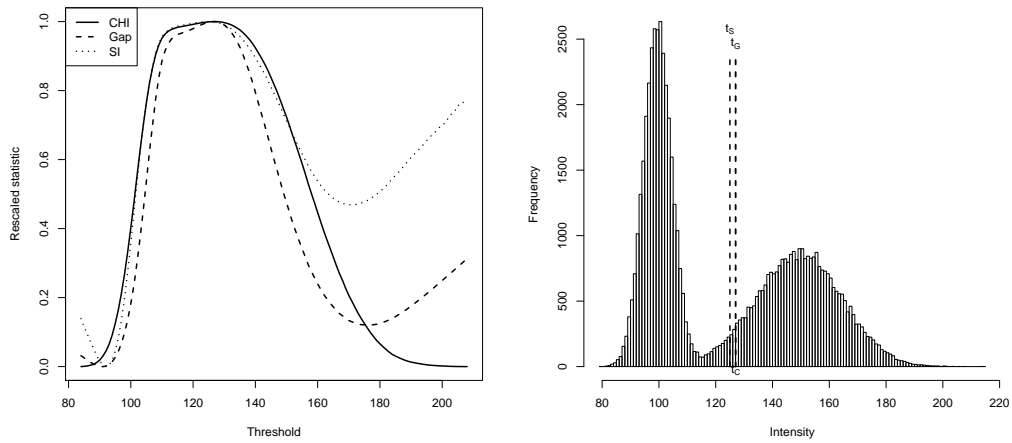


Figure 2: For \mathcal{X}_2 . Caption is as for \mathcal{X}_1 in Figure 1.

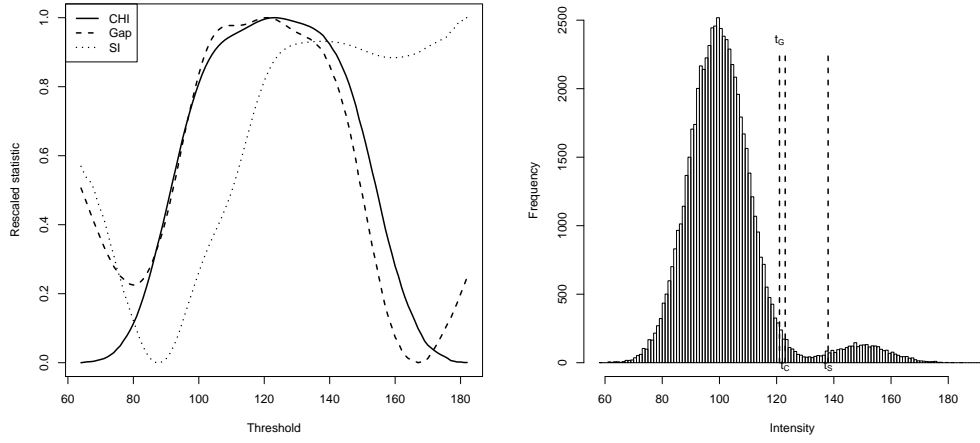


Figure 3: For \mathcal{X}_3 . Caption is as for \mathcal{X}_1 in Figure 1.

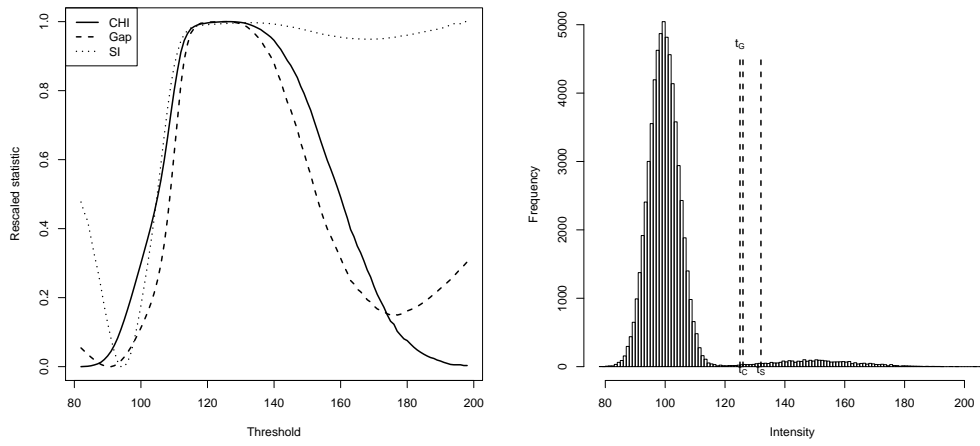


Figure 4: For \mathcal{X}_4 . Caption is as for \mathcal{X}_1 in Figure 1.

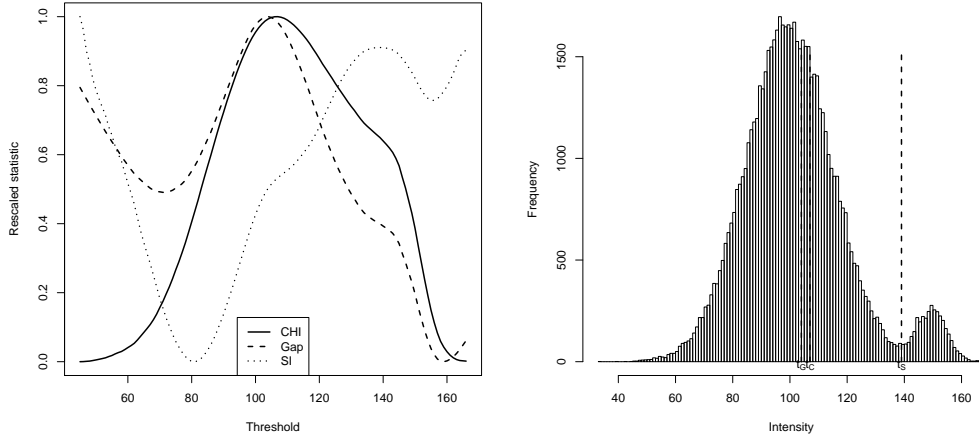


Figure 5: For \mathcal{X}_5 . Caption is as for \mathcal{X}_1 in Figure 1.

219 Equation (13) shows that the SI-based method prefers larger $b(x, t)$ and
 220 smaller $a(x, t)$ for the intensities x of the majority group. Hence, when the
 221 two group sizes are very different from each other, the SI-based method may
 222 favour an even more extreme partition to increase $b(x, t)$ for the majority
 223 group, in particular when the minority group has a flat and wide histogram
 224 (i.e., a large variance of its intensity distribution). In this case, as shown in
 225 Figures 3-5, a global maximum of $SI(t)$ may occur at an end of the intensity
 226 range; to avoid selecting such a maximum as t^* , we search for the local
 227 maximum next to it. We define a point to be a local maximum if its value is
 228 no less than those of three intensities in its nearest neighbourhood on both
 229 sides; this characterisation works empirically in our experiments.

230 The histograms constructed from the five simulated data sets are plotted
 231 in the right-hand panels of Figures 1-5, with corresponding thresholds t_C^* , t_G^*
 232 and t_S^* superimposed.

233 From these plots, we can observe that, for data sets \mathcal{X}_1 - \mathcal{X}_4 , the three meth-
 234 ods perform similarly well; in other words, the thresholding performances of
 235 the Gap-statistic-based method and the SI-based method are comparable to
 236 that of the popular Otsu's method. For \mathcal{X}_5 , the SI-based method performs
 237 the best; one explanation of why Otsu's method performs poorly in the case is
 238 that, as mentioned in section 2.2, the CHI is simply an F statistic in ANOVA
 239 or Student's t statistic for two-group comparison, which is not so reliable if

240 the two groups are of remarkably different group sizes and within-group vari-
 241 ances. More discussion about this can be found in Xue and Titterington
 242 (2010).

243 3.2. Real images

244 Here we choose five real, publicly-available images to demonstrate the
 245 performance of the three binarisation methods, which are based on equations
 246 (5), (9) and (13) by using $\text{CHI}(t)$, $\text{Gap}(t)$ and $\text{SI}(t)$, respectively.

247 The five images are ‘Bonemarr’, ‘Cameraman’, ‘Coins’, ‘NDT-image20’
 248 and ‘NDT-image5’. Two reasons why we choose these five images are as
 249 follows. First, ‘Bonemarr’, ‘Cameraman’ and ‘Coins’, were used by the
 250 MATLAB Image Processing Toolbox for demonstration or by the image-
 251 thresholding literature including Sahoo et al. (1988), Sahoo and Arora (2004)
 252 and Sahoo and Arora (2006), and ‘NDT-image20’ and ‘NDT-image5’ are two
 253 nondestructive testing images used in a recent comprehensive survey (Sezgin
 254 and Sankur, 2004) of image-binarisation methods. Secondly, these images
 255 represent roughly four types of image that we mentioned in section 3.1 and
 256 one case in which all three methods fail to provide an acceptable threshold,
 257 as shown shortly.

258 In Figures 6-10, for each of the five images, we show its original version,
 259 the curves for its statistics of $\text{CHI}(t)$, $\text{Gap}(t)$ and $\text{SI}(t)$, its histogram with
 260 corresponding thresholds t_C^* , t_G^* and t_S^* superimposed and its binarised ver-
 261 sions based on t_C^* , t_G^* and t_S^* , respectively. The values of t_C^* , t_G^* and t_S^* are
 262 listed in Table 2.

	t_C^*	t_G^*	t_S^*
Bonemarr	154	182	155
Cameraman	89	44	81
Coins	127	80	125
NDT-image20	80	39	69
NDT-image5	150	234	85

Table 2: Thresholds t_C^* , t_G^* and t_S^* selected by the use of the CHI, the Gap statistic and the overall average Silhouette width, respectively, for five real images.

263 Suppose that we are interested in separating large white cells (as \mathcal{C}_2) from
 264 other material (as \mathcal{C}_1) in ‘Bonemarr’, from Figure 6. We can observe that \mathcal{C}_1
 265 and \mathcal{C}_2 are of similar sizes and different spreads. In this sense, ‘Bonemarr’ is

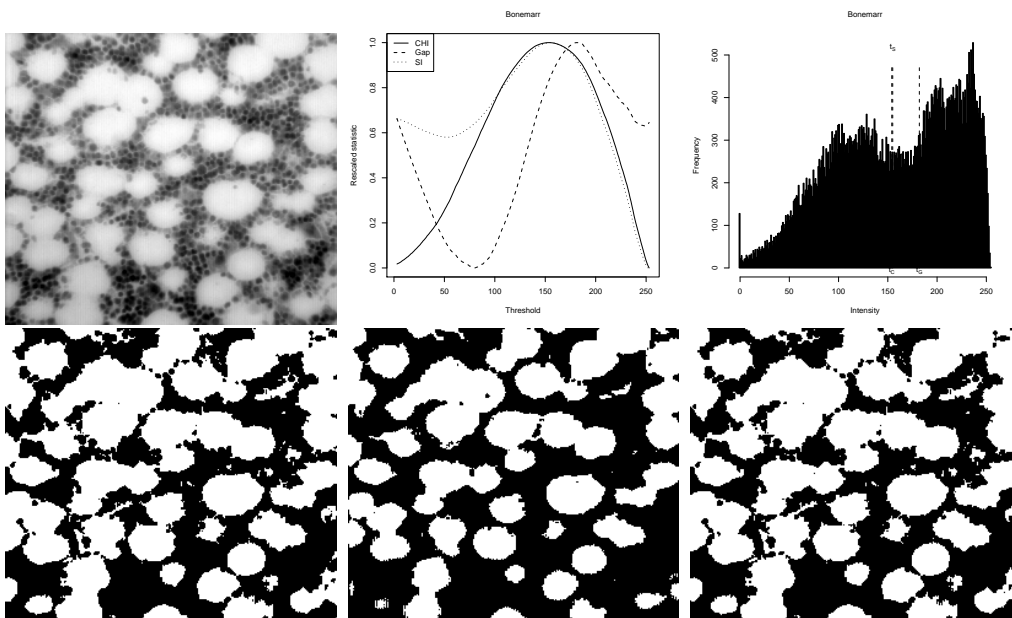


Figure 6: For ‘Bonemarr’. Upper row, from left to right: original image; $CHI(t)$, $Gap(t)$ and $SI(t)$, all rescaled to the range $[0, 1]$; histogram with thresholds t_C^* , t_G^* and t_S^* indicated by dashed lines. Lower row, from left to right: binarised images by using t_C^* , t_G^* and t_S^* , respectively.

266 similar to \mathcal{X}_2 in section 3.1, except that the two groups have much heavier
 267 overlap.

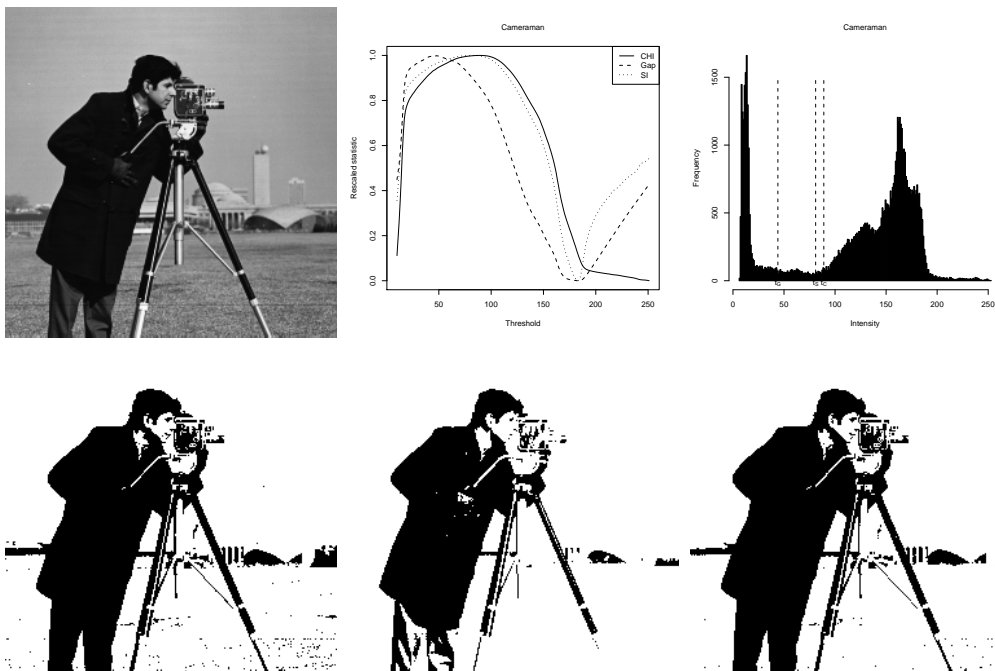


Figure 7: For ‘Cameraman’. Caption is as for Figure 6.

268 The ‘Cameraman’ image has been widely used as a benchmark for bi-
 269 narisiation, although its histogram can be regarded as multi-modal to some
 270 extent. If we assume that the cameraman and the camera in the image com-
 271 pose the foreground \mathcal{C}_1 and the rest composes the background \mathcal{C}_2 , the two
 272 groups are of quite different sizes whereas their spreads are similar (note the
 273 long tail of the histogram of \mathcal{C}_1 as shown in Figure 7). This makes ‘Came-
 274 raman’ similar to \mathcal{X}_3 .

275 As shown in Figure 8, the histogram of ‘Coins’ is similar to that of \mathcal{X}_4 ,
 276 in the sense that the dark background (\mathcal{C}_1) contains many more pixels, and
 277 has much more compact intensities, than the gray coins (\mathcal{C}_2). In contrast,
 278 in the histogram of a light microscope image ‘NDT-image20’ of a material
 279 structure as shown in Figure 9, the majority group of particles (\mathcal{C}_2 here) has
 280 more variable intensities than the minority group of pores (\mathcal{C}_1), and thus the
 281 histogram is similar to that of \mathcal{X}_5 .

282 In the four cases of ‘Bonemarr’, ‘Cameraman’, ‘Coins’ and ‘NDT-image20’,

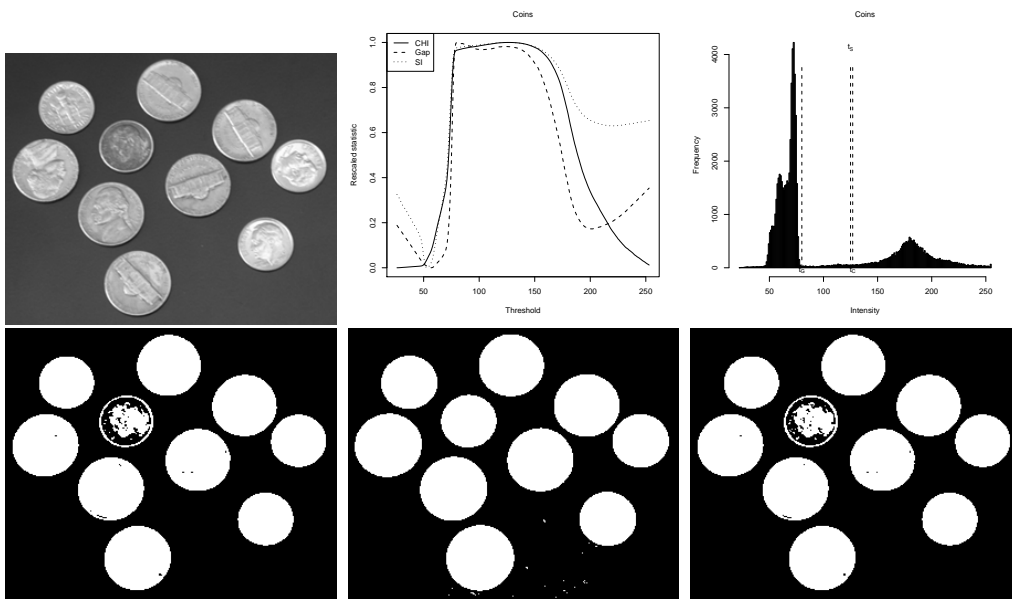


Figure 8: For 'Coins'. Caption is as for Figure 6.

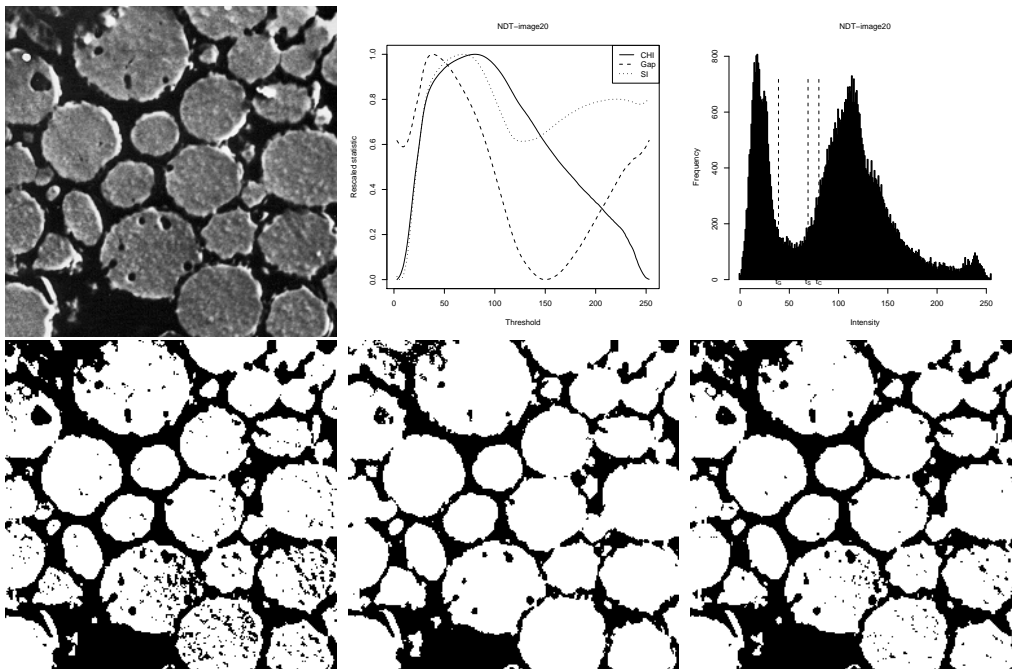


Figure 9: For 'NDT-image20'. Caption is as for Figure 6.

283 t_C^* and t_S^* are similar whereas t_G^* is further away from the middle of the in-
 284 tensity range, as shown in Table 2 and Figures 6-9. All thresholds, t_C^* , t_G^*
 285 and t_S^* , are located between the histogram modes for \mathcal{C}_1 and \mathcal{C}_2 , and produce
 286 similar, reasonable binarisation results, although using t_G^* in general provides
 287 better results, in particular for ‘Coins’ and ‘NDT-image20’.

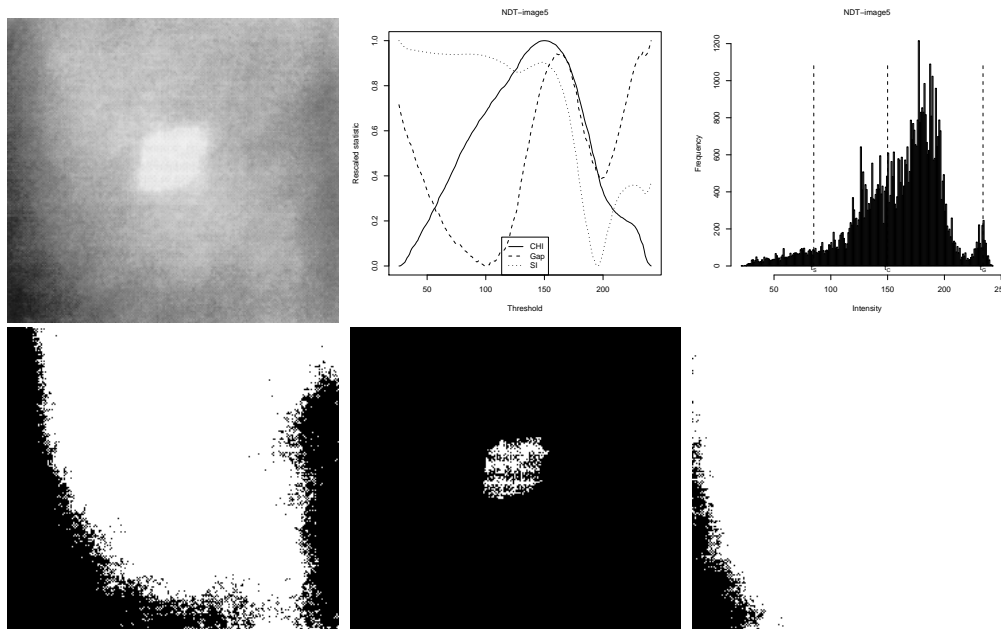


Figure 10: For ‘NDT-image5’. Caption is as for Figure 6.

288 Here we also use ‘NDT-image5’, an infrared thermal image, to present an
 289 example for which all three methods fail to find a reasonable threshold. From
 290 Figure 10, we can observe a histogram very similar to that of \mathcal{X}_5 , where the
 291 minority group of a bright hole (\mathcal{C}_2 here) located in the centre of the image is
 292 of extremely small size and small variance. In this case, the values of t_C^* and t_S^*
 293 are too small, while that of t_G^* is too large, to extract the hole. Nevertheless,
 294 an interesting phenomenon, that can be observed from the curve of $SI(t)$ in
 295 Figure10, is that its rightmost local maximum corresponds to a threshold
 296 that can extract the hole very well, and its second rightmost local maximum
 297 corresponds to t_C^* , the threshold obtained by Otsu’s method.

298 4. Conclusions

299 It was our intention in this paper to demonstrate that some statistics
300 originally developed in cluster analysis for determining K , the number of
301 clusters, can instead be appropriately used in image thresholding to select
302 \mathbf{t}^* , the optimal thresholds. We chose three highly-cited statistics, namely
303 the CHI, the Gap statistic and the Silhouette widths, and applied them
304 first to simulated data and then to real images. The CHI-based method
305 is equivalent to Otsu's method; the other two methods provide comparable
306 performance to that of Otsu's method, which is encouraging. In the end, we
307 would emphasise that, instead of developing or recommending a particular
308 image-thresholding method, which is certainly important, it was more our
309 expectation that some investigation into using many other K -determination
310 statistics for image thresholding, or more general image segmentation, can
311 be encouraged by the results presented in this paper.

312 Acknowledgements

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

317 References

- 318 Chiang, M. M.-T., Mirkin, B., 2010. Intelligent choice of the number of clus-
319 ters in K-means clustering: an experimental study with different cluster
320 spreads. *Journal of Classification* 27 (1), 3–40.
- 321 Glasbey, C. A., 1993. An analysis of histogram-based thresholding algo-
322 rithms. *CVGIP: Graphical Models and Image Processing* 55 (6), 532–537.
- 323 Jain, A. K., 2010. Data clustering: 50 years beyond K-means. *Pattern Recog-
324 nition Letters* 31 (8), 651–666.
- 325 Jain, A. K., Murty, M. N., Flynn, P. J., 1999. Data clustering: a review.
326 *ACM Computing Surveys* 31 (3), 264–323.
- 327 Kurita, T., Otsu, N., Abdelmalek, N., 1992. Maximum likelihood thresh-
328 olding based on population mixture models. *Pattern Recognition* 25 (10),
329 1231–1240.

- 330 Milligan, G. W., Cooper, M. C., 1985. An examination of procedures for
331 determining the number of clusters in a data set. *Psychometrika* 50 (2),
332 159–179.
- 333 Otsu, N., 1979. A threshold selection method from gray-level histograms.
334 *IEEE Transactions on Systems, Man, and Cybernetics* SMC-9, 62–66.
- 335 Rousseeuw, P. J., 1987. Silhouettes: a graphical aid to the interpretation
336 and validation of cluster analysis. *Journal of Computational and Applied*
337 *Mathematics* 20, 53–65.
- 338 Sahoo, P. K., Arora, G., 2004. A thresholding method based on two-
339 dimensional Renyi’s entropy. *Pattern Recognition* 37 (6), 1149–1161.
- 340 Sahoo, P. K., Arora, G., 2006. Image thresholding using two-dimensional
341 Tsallis-Havrda-Charvát entropy. *Pattern Recognition Letters* 27 (6), 520–
342 528.
- 343 Sahoo, P. K., Soltani, S., Wong, A. K. C., Chen, Y. C., 1988. A survey of
344 thresholding techniques. *Computer Vision, Graphics, and Image Process-*
345 *ing* 41 (2), 233–260.
- 346 Sezgin, M., Sankur, B., 2004. Survey over image thresholding techniques and
347 quantitative performance evaluation. *Journal of Electronic Imaging* 13 (1),
348 146–165.
- 349 Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of
350 clusters in a dataset via the gap statistic. *Journal of the Royal Statistical*
351 *Society, Series B* 63 (2), 411–423.
- 352 Trier, Ø. D., Jain, A. K., 1995. Goal-directed evaluation of binarization
353 methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*
354 17 (12), 1191–1201.
- 355 Xu, R., Wunsch, D., 2005. Survey of clustering algorithms. *IEEE Transac-*
356 *tions on Neural Network* 16 (3), 645–678.
- 357 Xue, J.-H., Titterington, D. M., 2010. *t*-tests, *F*-tests and Otsu’s methods
358 for image thresholding. manuscript.