OPTIMAL GLOBAL THRESHOLD ESTIMATION USING STATISTICAL CHANGE-POINT DETECTION

Rohit Kamal Chatterjee[1] and Avijit Kar[2]

[1]Department of Computer Science & Engineering,
Birla Institute of Technology, Mesra, Ranchi, India.
[2]Department of Computer Science & Engineering,
Jadavpur University, Kolkata, India.

ABSTRACT

Aim of this paper is reformulation of global image thresholding problem as a well-founded statistical method known as change-point detection (CPD) problem. Our proposed CPD thresholding algorithm does not assume any prior statistical distribution of background and object grey levels. Further, this method is less influenced by an outlier due to our judicious derivation of a robust criterion function depending on Kullback-Leibler (KL) divergence measure. Experimental result shows efficacy of proposed method compared to other popular methods available for global image thresholding. In this paper we also propose a performance criterion for comparison of thresholding algorithms. This performance criteria does not depend on any ground truth image. We have used this performance criterion to compare the results of proposed thresholding algorithm with most cited global thresholding algorithms in the literature.

KEYWORDS

Global image thresholding, Change-point detection, Kullback-Leibler divergence, robust statistical measure, thresholding performance criteria.

1. INTRODUCTION

A grey-level digital image is a two dimensional signal $I : \mathbb{Z}\times\mathbb{Z} \rightarrow L$, where $L=\{\ l_i \in \mathbb{R}$ and $i=1,2,\ldots,M\}$ is the set of M grey-levels. The problem of automatic thresholding is to estimate an optimal threshold t_0 which segments the image into two meaningful sets, viz. background $B=\{b_b(x,y)=1 \mid I(x,y)<t_0\}$ and foreground $F=\{b_f(x,y)=1 \mid I(x,y)\geq t_0\}$ or the opposite. The function $I(x,y)$ can take any random value $l_i \in L$; so, sampling distribution of grey levels becomes an important deciding factor for t_0. In many image processing applications, automating the process of optimal thresholding is extremely important for low-level segmentation or even final segmentation of object and background.

In general, automatic thresholding algorithms are divided into two groups, viz. global and local methods. Global methods estimate a single threshold for the entire image; local methods find an adaptive threshold for each pixel depending on the characteristics of its neighborhood. Global methods are used if the image is considered as a mixture of two or more statistical distributions. In this paper, we address the global thresholding methods guided by the image histogram. Most of the cases global thresholding methods try to estimate the threshold (t_0) iteratively by optimizing a criterion function [1]. Some other methods attempt to estimate optimal t_0 depending on histogram shape [2, 3], image attribute such as topology [5] or some clustering techniques [4, 8, 20]. Comprehensive surveys discussing various aspects of thresholding methods can be found in the references [1, 6, 7].

DOI : 10.5121/sipij.2017.8402

Many of these classical and recent schemes perform remarkably well for images with matching underlying assumptions but fail to yield desired results otherwise. Some of the explicit or implicit reasons for their failure could be: (i) assumption of some standard distribution (e.g. Gaussian) [19], in reality though, foreground and background classes can have arbitrary asymmetric distributions, (ii) use of non-robust measures for computing criterion functions which get influenced by outliers. Further, the effectiveness of these algorithms greatly decreases when the areas under the two classes are highly unbalanced. Some of the methods depend on user specified constant (e.g. Renyi or Tsallis entropy based methods) [17, 18], greatly compromising their performance without its appropriate value.

This paper proposes an algorithm for addressing these drawbacks using a statistical technique known as *change-point detection (CPD).* For the last few decades, models of change-point detection are successfully applied by researchers in statistics and control theory for detecting abrupt changes in the statistical behavior of an observed signal or time series [9]. The general principle of change-point detection considers an observed sequence of independent random variables $\{Y_k\}_{1\ldots n}$ with a probability density function (pdf) $p_\theta(y)$ depending on a parameter θ. If any change occurs in the sequence then it is assumed that parameter θ takes a value θ_0 before any change and at some unknown time $\mathbf{t_0}$ alters to θ_1 ($\neq\theta_0$). The main problem of statistical *change-point detection* is to decide the change in parameter and also the time of change. The theory of CPD is used in this paper to decide the global threshold in an image depending on the change in the histogram.

Further, in section 4 of this paper, we propose a new performance index for the evaluation of thresholding algorithms. It depends on the structural difference between the shapes of background and foreground. The advantage of this performance index is that it does not depend on any ground truth image. We use this performance index to compare different thresholding algorithms including ours.

Rest of the paper is organized as follows: Section 2 provides a short introduction to the problem of statistical change-point detection, section 3 formulates and derives the global thresholding as a change-point detection problem, section 4 describes our proposal for thresholding performance criteria, section 5 presents the experimental results and compares the results with various often cited global thresholding algorithms, and finally section 6 summarizes main ideas in this paper.

2. THE CHANGE-POINT DETECTION (CPD) PROBLEM

The Change-point detection (CPD) problem can be classified into two broad categories: *real-time* or *online* and *retrospective* or *offline* change-point detection. The first targets applications where the instantaneous response is desired such as robot control; on the other hand, *retrospective* or *offline* change-point detection is used when longer reaction periods are allowed e.g. image processing problems [10]. The later technique is likely to give more accurate detection since the entire sample distribution is accessible. Since the image and the corresponding histogram are available to us, we concentrate on offline change-point detection in this paper. We also assume that there is only one change point throughout the given observations $\{y_k\}_{1\ldots n}$. When required, this assumption can easily be relaxed and extended to multiple change point detection that can be applied in multi-level threshold detection problems.

2.1. Problem Statement

When taking an offline point of view about the observations $\mathbf{y_1, y_2 \ldots, y_n}$ with corresponding probability distribution functions $\boldsymbol{F_1, F_2, \ldots, F_n}$, belong to a common parametric family $\boldsymbol{F(\varphi)}$, where $\boldsymbol{\varphi \in R^p}$, p>0. Then the change point problem is to test the null hypothesis (H_0) about the population parameter $\boldsymbol{\varphi_j}$, $j = 1,2, \ldots, n$:

$$H_0 : \phi_j = \theta_0 \quad for \quad i \le j \le n$$

versus an alternative hypothesis (1)

$$H_1 : \phi_j = \begin{cases} \theta_0, & for \quad 1 \le j \le k \\ \theta_1, & for \quad k < j \le n \end{cases}$$

where $\theta_0 \neq \theta_1$ and k is an unknown time of change.

These hypotheses together disclose the characteristics of change point inference, determining if any change point exists in the process and estimating the *time of change* $t_0 = k$. The *likelihood ratio* corresponding to the hypotheses H_0 and H_1 is given by

$$\Lambda_1^n(k) = \frac{\prod_{j=1}^{k-1} p_{\theta_0}(y_j) \times \prod_{j=k}^{n} p_{\theta_1}(y_j)}{\prod_{j=1}^{n} p_{\vartheta}(y_j)} \tag{2}$$

where p_{θ_0} and p_{θ_1} are pdfs before and after the change occurs and p_{ϑ} is the overall probability density. When the only unknown parameter is t_0, its maximum likelihood estimate (MLE) is given by the following statistic

$$t_0 = arg \max_{1<k<n} \Lambda_1^n(k) \tag{3}$$

2.2. Offline Estimation of the Change Time

When the problem is to estimate the change time (t_0) in the sequence of observations $\{y_j\}_{1\ldots n}$ and if we assume the existence of a change point with the same presumption as in the last section. Therefore, considering equation (2) and (3) and the fact that $\prod_1^n p_{\vartheta}$ is a constant for a given data, the corresponding MLE estimate is

$$\hat{t}_0 = \arg \max_{1 \le k \le n} \ln\left(\prod_{j=1}^{k-1} p_{\theta_0}(y_j) \times \prod_{j=k}^{n} p_{\theta_1}(y_j)\right) \tag{4}$$

where $\hat{t}_0$ is a maximum log-likelihood estimate of t_0. Rewriting equation (4) as

$$\hat{t}_0 = \arg \max_{1 \le k \le n} \left[ln\left(\frac{\prod_{j=k}^{n} p_{\theta_1}(y_j)}{\prod_{j=k}^{n} p_{\theta_0}(y_j)}\right) + ln\left(\prod_{j=1}^{n} p_{\theta}(y_j)\right)\right] \tag{5}$$

As $ln(\prod_{j=1}^{n} p_{\theta}(y_j))$ remains constant for a given observation, estimation of $\hat{t}_0$ is simplified as

$$\hat{t}_0 = \arg \max_{1 \le k \le n} \sum_{j=k}^{n} ln\left(\frac{p_{\theta_1}(y_j)}{p_{\theta_0}(y_j)}\right) \tag{6}$$

Therefore, the MLE of the change time $\mathbf{t_0}$ is the value which maximizes the sum of log-likelihood ratio corresponding to all k possible values given by equation (6).

3. CHANGE-POINT DETECTION FORMULATION OF GLOBAL THRESHOLDING

3.1. Assumptions

Let $(\chi, \beta_\chi, P_\theta)_{\theta\in\Theta}$ be the *statistical space of discrete grey-levels* associated with a random variable $Y: \mathbb{Z}x\mathbb{Z} \rightarrow \mathbb{Z}$, where β_χ is the σ-field of Borel subsets $\mathbf{A} \subset \chi$ and $\{P_\theta\}_{\theta\in\Theta}$ is a family of probability distributions defined on the measurable space (χ, β_χ) with parameter space Θ, an open subset of

$\mathbb{R}^q$, q>0.We consider a finite population $\boldsymbol{\Pi}$ of all gray-level images with $\boldsymbol{N}$ elements that could be classified into $\boldsymbol{M}$ categories or classes $\boldsymbol{L}=\{\boldsymbol{l_1}, ..., \boldsymbol{l_M}\}$, i.e. each sample point in the sample image can take any random gray-level values from the set $\boldsymbol{L}$.

3.2. Change-Point Detection Formulation

Since we are mainly interested in discrete gray-level data, we consider the *multinomial distribution* model. Let $\wp=\{E_i\}$, i=1,...,M be a partition of χ. The formula $\boldsymbol{Pr_\theta(E_i) = p_i(\theta)}$, $i = 1, . . ., M$, defines the probability of the $\boldsymbol{l_i}^{th}$ gray-level in the discrete statistical model. Further we assume $\{y_1, . . ., y_N\}$ to be a random sample from the population described by the random variable $\mathbf{Y}$, representing the gray-level of a pixel. And let $\boldsymbol{N_i = \sum_{j=1}^{N} I_{E_i}(y_j)}$, where I_E is the index function. Then we can approximate $\boldsymbol{p_i}(\theta)\boldsymbol{\approx N_i/N}$, $i=1,\dots, M$. Estimating θ by maximum likelihood method consists of maximizing the joint probability distribution for fixed $n_1, . . . , n_M$,

$$\boldsymbol{Pr_\theta(N_1 = n_1, N_2 = n_2 \dots, N_M = n_M) = \frac{n!}{n_1!\, n_2! \dots n_M!}(p_1(\theta))^{n_1}(p_2(\theta))^{n_2} \dots (p_M(\theta))^{n_M}} \tag{7}$$

or equivalently maximizing the log-likelihood function

$$\boldsymbol{\Lambda(\theta) = ln\left[\frac{n!}{n_1!\, n_2! \dots n_M!}(p_1(\theta))^{n_1}(p_2(\theta))^{n_2} \dots (p_M(\theta))^{n_M}\right]} \tag{8}$$

Therefore, referring to equation (4), problem of estimating the threshold by MLE can be stated as

$$\boldsymbol{\hat{t}_0 = \quad arg \max_{1\le j\le M} \ln\left(\frac{n!}{n_1!\, n_2! \dots n_M!}\prod_{i=1}^{j-1}(p_i(\theta_0))^{n_i}\prod_{i=j}^{M}(p_i(\theta_1))^{n_i}\right)} \tag{9}$$

where unknown parameter $\boldsymbol{\theta = \theta_0}$ before the change and $\boldsymbol{\theta = \theta_1}$ after the change. Now, equation (9) can be expanded as

$$\boldsymbol{\hat{t}_0 = arg \max_{1\le j\le M}[ln\left(\frac{n!}{n_1!\, n_2! \dots n_M!}\right) + ln\left(\frac{\prod_{i=j}^{M}(p_i(\theta_1))^{n_i}}{\prod_{i=j}^{M}(p_i(\theta_0))^{n_i}}\right) + ln\left(\prod_{i=1}^{M}(p_i(\theta_0))^{n_i}\right)]} \tag{10}$$

The first term within the bracket on the right side of equation (10) is a constant and the last term is independent of j, i.e. it cannot influence the MLE. So, eliminating these terms from equation (10) and simplifying we get

$$\boldsymbol{\hat{t}_0 = arg \max_{1\le j\le M}\sum_{i=j}^{M} ln\left(\frac{p_i(\theta_1)}{p_i(\theta_0)}\right)^{n_i}} \tag{11}$$

Multiplying and dividing $\boldsymbol{N}$ on right side of equation (11) we get

$$\boldsymbol{\hat{t}_0 = arg \max_{1\le j\le M} N\sum_{i=j}^{M}\left(\frac{n_i}{N}\right) ln\left(\frac{p_i(\theta_1)}{p_i(\theta_0)}\right)} \tag{12}$$

assuming $\boldsymbol{p_i(\theta)\approx n_i/N}$ equation (12) can be written as

$$\hat{t}_0 \approx arg \max_{1\leq j\leq M} N \sum_{i=j}^{M} p_i(\theta_1) ln\left(\frac{p_i(\theta_1)}{p_i(\theta_0)}\right) \tag{13}$$

The expression in (13) under the summation denotes *Kullback-Leibler (KL)* divergence between the density $p(\theta_1)$ and $p(\theta_0)$, where $p(\theta_1)$ and $p(\theta_0)$ denotes the pdfs above and below the threshold location j; therefore equation (13) can be written as

$$\hat{t}_0 = arg \max_{1\leq j\leq M} N K_j^M (p(\theta_1)\,||p(\theta_0)) \tag{14}$$

Since total sum $\sum_{j=1}^{M} p_j(\theta_1) ln\left(\frac{p_j(\theta_1)}{p_j(\theta_0)}\right)$ is independent of j, i.e. a constant for a given observation, a sample image, therefore equation (14) can be rewritten as

$$\hat{t}_0 = arg \min_{1\leq j\leq M} N K_1^j (p(\theta_1)||p(\theta_0)) \tag{15}$$

Hence, equation (15) provides the maximum likelihood estimation of the threshold t_0. Equation (15) can be restated as the following proposition:

Proposition 1: *In a mixture of distributions, the maximum likelihood estimate of change-point is found by minimizing the Kullback-Leibler divergence of the probability mass across successive thresholds.*

In spite of this striking property, KL divergence is not a 'metric' since it is not symmetric. An alternative symmetric formula by "averaging" the two KL divergences is given as [11]

$$D(p_{\theta_1}||p_{\theta_0}) = \frac{1}{2}\left(K(p_{\theta_1}||p_{\theta_0}) + K(p_{\theta_0}||p_{\theta_1})\right) \tag{16}$$

An attractive property of KL divergence is its robustness i.e. KL divergence is little influenced even when one component of mixture distribution is considerably skewed. A proof of robustness can be found for generalized divergence measures in [11, 12].

This method can be easily extended to find multiple thresholds for several mixture distributions by identifying multiple change-points simultaneously.

3.3. Implementation

Section Let us consider an image $\boldsymbol{I}: \mathbb{Z}x\mathbb{Z} \rightarrow \boldsymbol{L}$, whose pixels assume M gray-levels in the set $\boldsymbol{L}=\{l_1, l_2,\ldots, l_M\}$. The empirical distribution of the image can be represented by a normalized histogram $\boldsymbol{p(l_i)=n_i/N}$, where $\boldsymbol{n_i}$ is the number of pixels in ith gray-level and $\boldsymbol{N}$ is the total number pixels in the image.

Now, suppose we are grouping the pixels into two classes **B** and **F** (background and object) by thresholding at the level $\boldsymbol{k}$. Histogram of gray-levels can be found for the classes **B** and **F**; let us denote them as $\boldsymbol{p_B(l_i)}$and $\boldsymbol{p_F(l_i)}$. Following statistics are calculated for the level $\boldsymbol{k}$.

$$\mathbf{K}^B(\boldsymbol{k}) = \sum_{i=1}^{k-1} \boldsymbol{p_F}(\mathbf{l}_i) \ln\left(\frac{p_F(\mathbf{l}_i)}{p_B(\mathbf{l}_i)}\right) \tag{17}$$

$$\mathbf{K}^F(\boldsymbol{k}) = \sum_{i=k}^{M} \boldsymbol{p_B}(\mathbf{l}_i) \ln\left(\frac{p_B(\mathbf{l}_i)}{p_F(\mathbf{l}_i)}\right) \tag{18}$$

and finally,

$$CPD(\mathbf{k}) = \frac{1}{2}\left(\mathrm{K}^{F}(k) + \mathrm{K}^{B}(k)\right) \tag{19}$$

The minimum value of ***CPD(k)*** for all values of ***k*** in the range [1,…, M] gives an optimal estimate of threshold t_0**.**

4. THRESHOLDING PERFORMANCE CRITERIA

The objective of the global thresholding algorithm is to divide the image into two binary images generally called background and foreground (object). Most of the histogram-based thresholding algorithms try to devise a criterion function which produces a threshold to separate the shapes and patterns of the foreground and background as much as possible. A good thresholding algorithm can be judged by how well it sets apart the object and the background binary images, i.e. how much dissimilarity exists between the foreground and the background. Since the background and foreground images are binary images dissimilarity between them can be measured by any binary distance measures. Based on this observation, we propose a threshold evaluation criterion, which tries to find the dissimilarity between the patterns and shapes in foreground and background.

A number of binary similarity and distance measures have been proposed in different areas, a comprehensive survey of them can be found in Choi et al. [13]. In order to understand the distance measure used in our work, it is helpful to refer to the following contingency table (**Table 1**):

Table 1. Binary contingency table

Foreground	Background 1	0
1	*a*	*b*
0	*c*	*d*

The cell entries in **Table 1** are the number of pixel locations for which the two binary images agree or differ. For example, cell entry '*a*' is the total count of pixel locations where both binary images take a value one. Hence, $b + c$ denote the total count where foreground and background pixels differ (Hamming distance) and $a+d$ is the total count where they agree.

In order to extract the shapes and patterns present in the foreground (**F**) and background (**B**) images, we use *binary morphological gradient*. The binary Morphological gradient is the difference between the *eroded* and *dilated* images. Obviously, any other edge or texture detection algorithm for binary images can be also used to extract the objects present in foreground and background.

In this paper, we use a simple binary distance measure known as *Normalised Manhattan distance* (D_{NM}) given by

$$D_{NM}\left(F_g, B_g\right) = \frac{b+c}{a+b+c+d} \tag{20}$$

where $\mathbf{F_g}$ and $\mathbf{B_g}$ denote Binary Morphological gradients of foreground (**F**) and background(**B**) respectively. The range of this distance measure is the interval *[0, 1]*. It is expected that well-segmented image will have D_{NM} close to **1,** while in the worst case D_{NM} =**0**. The advantage of this algorithm is that it does not require any ground truth image.

5. EXPERIMENTAL RESULTS WITH DISCUSSION

To validate the applicability of proposed Change-Point Detection (CPD) thresholding algorithm, we provide experimental results and compare the results with existing algorithms. The first row of **Figure 1** shows test images that are labeled from left to right as *Dice, Rice, Object, Denise, Train,* and *Lena* respectively.

TABLE 2: Threshold evaluation criterion (D_{NM}) for the test images (A) Dice, (B) Rice, (C) Object, (D) Denise, (E) Train, and (F) Lena

Dice	Threshold	D_{NM}
Entropy	157	0.1046
Kittler	55	0.1584
Kurita	106	0.1574
Otsu	103	0.1534
Sahoo	134	0.1274
CPD	96	0.2746

(A)

Rice	Threshold	D_{NM}
Entropy	118	0.2360
Kittler	132	0.2210
Kurita	126	0.2286
Otsu	124	0.2306
Sahoo	133	0.2190
CPD	107	0.2458

(B)

Object	Threshold	D_{NM}
Entropy	150	0.1320
Kittler	83	0.0558
Kurita	107	0.0634
Otsu	110	0.0652
Sahoo	154	0.1550
CPD	190	0.2944

(C)

Denise	Threshold	D_{NM}
Entropy	131	0.1768
Kittler	235	0.0426
Kurita	151	0.1612
Otsu	147	0.1650
Sahoo	128	0.1810
CPD	111	0.2078

(D)

Train	Threshold	D_{NM}
Entropy	117	0.4186
Kittler	234	0.0554
Kurita	150	0.3226
Otsu	142	0.3554
Sahoo	80	0.4564
CPD	108	0.4690

(E)

Lena	Threshold	D_{NM}
Entropy	118	0.1090
Kittler	102	0.1278
Kurita	89	0.1714
Otsu	72	0.1874
Sahoo	109	0.1120
CPD	64	0.2056

(F)

The images have deliberately been so selected that the difference of areas between foreground and background is hugely disproportionate. This gives us an opportunity to test the robustness of CPD algorithm. To compare the results, we selected five most popular thresholding algorithms, namely, Kittler-Illingworth [14], Otsu [15], Kurita [16], Sahoo [17] and Entropy [18].

In **Figure 1** third row onwards show the outputs of different thresholding algorithms. The last row shows the output of the proposed CPD thresholding algorithm. Due to substantial skewness in the distributions of gray-levels in object or background, most of the algorithms confused foreground with background. But results in the last row clearly show that CPD works significantly better in all cases.

Table 2 shows optimal thresholds of five selected algorithms and the CPD algorithm using our proposed performance criteria. It is clear that CPD performs reasonably well. For example, consider the Denise image and Train image, Kittler-Illingworth thresholding totally fails to distinguish the object from the background due to its assumption of Gaussian distribution for both foreground and background [19]. Otsu's and Kurita's method yield almost same output due to their common assumptions. Corresponding histograms are also reproduced in **Figure 2** marked with threshold locations of all the six algorithms above for reference. The threshold locations show that CPD algorithm is very little influenced by the asymmetry of object of background distributions.

6. CONCLUSIONS

In this paper we propose a novel global image thresholding algorithm based on Statistical Change-Point detection (CPD), which is derived based on a symmetric version of Kullback-Leibler divergence measure. The experimental results clearly show this algorithm is largely unaffected by disproportionate dispersal of object and background scene and also very little influenced by the skewness of distributions of object and background compared to other well-known algorithms. We also propose a thresholding performance criterion using dissimilarity between foreground and background binary images. Advantage of this performance criterion is that it does not require any ground truth image.

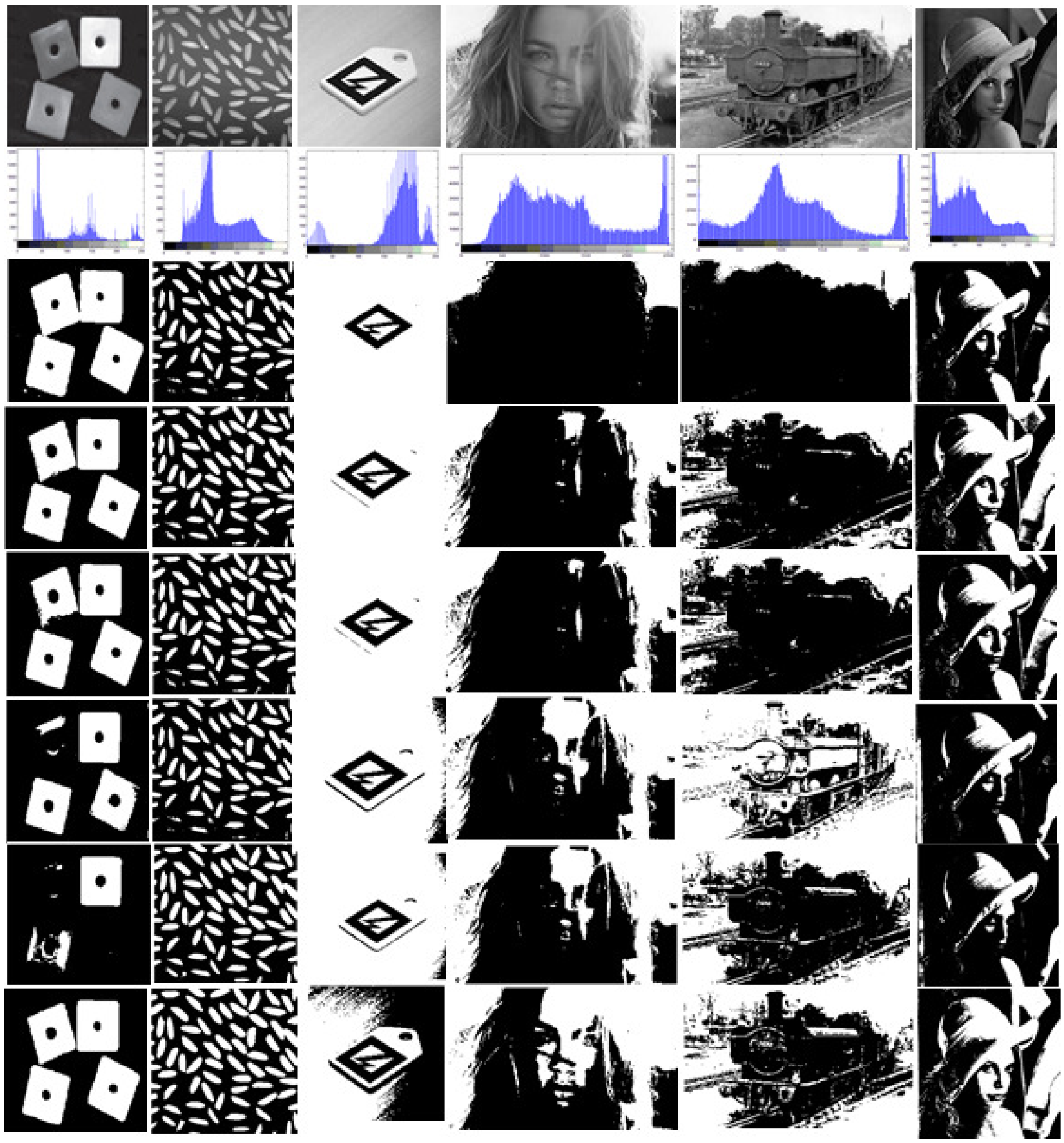

Figure 1. Result of thresholding algorithms on tested images: **Row-1**: Original Images; **Row-2**: Shapes of histograms; **Row-3**: Kittler; **Row-4**: Otsu; **Row-5**: Kurita; **Row-6**: Sahoo; **Row-7**: Entropy: **Row-8**: CPD Threshold.

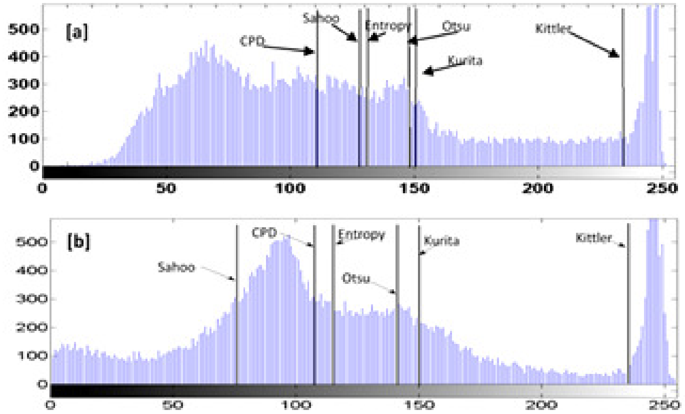

Figure 2: Histogram of (a) Denise and (b) Train image with threshold locations

REFERENCES

[1] M. Sezgin and B. Sankur (2004) "Survey over image thresholding techniques and quantitative performance evaluation", J. of Electronic Imaging, Vol. 13, No. 1, pp.146–165.

[2] A. Rosenfeld and P. De la Torre, (1983) "Histogram concavity analysis as an aid in threshold selection", IEEE Trans. Syst. Man Cybernetics. SMC-13, pp. 231–235.

[3] M. I. Sezan, (1985) "A peak detection algorithm and its application to histogram-based image data reduction", Graph. Models Image Process. Vol. 29, pp.47–59.

[4] D. M. Tsai, (1995) "A fast thresholding selection procedure for multimodal and unimodal histograms", Pattern Recogn. Lett. Vol. 16, pp. 653–666.

[5] A. Pikaz and A. Averbuch, (1996) "Digital image thresholding based on topological stable state", Pattern Recogn. Vol. 29, pp.829–843.

[6] N. R. Pal and S. K. Pal, (1993) "A review on image segmentation techniques," Pattern Recog., vol. 26, no. 9, pp. 1277–1294.

[7] P. K. Sahoo, S. Soltani, A. K. C. Wong, and Y. C. Chen, (1988) "A survey of thresholding techniques," Computer Vision, Graphics, and Image Process., vol. 41, no. 2, pp.233–260.

[8] C. V. Jawahar, P. K. Biswas, and A. K. Ray, (1997) ''Investigations on fuzzy thresholding based on fuzzy clustering,'' Pattern Recogn., vol. 30, no. 10, pp. 1605–1613.

[9] H. V. Poor, O. Hadjiliadis, Quickest Detection, Cambridge University Press, New York, 2009.

[10] J. Chen, A. K. Gupta, (2012) Parametric statistical change point analysis, with applications to genetics, medicine and finance, 2nd Ed., Birkhäuser, Boston.

[11] Y. Wang, (2011) "Generalized Information Theory: A Review and Outlook", J. of Inform. Tech., Vol. 10, No. 3, pp. 461-469.

[12] L. Pardo, (2006) Statistical Inference Based on Divergence Measures, Chapman & Hall/CRC, pp. 233.

[13] S. Choi, S. Cha, C. C. Tappert, (2010) "A Survey of Binary Similarity and Distance Measures," Systemics, Cybernetics and Informatics Vol. 8, No. 1, pp. 43-48, 2010.

[14] J. Kittler and J. Illingworth (1986) “Minimum error thresholding”, Pattern Recognition, Vol. 19, pp. 41–47.

[15] N. Otsu, (1979) “A threshold selection method from gray level histograms”, IEEE Trans. Syst. Man Cybern. SMC-9, pp. 62–66.

[16] T. Kurita, N. Otsu, and N. Abdelmalek, (1992) “Maximum likelihood thresholding based on population mixture models”. Pattern Recognition, Vol. 25, pp. 1231-1240.

[17] P. Sahoo, C. Wilkins, and J. Yeager, (1997) “Threshold selection using Renyi’s entropy”, Pattern Recogn. Vol. 30, pp. 71–84.

[18] P.K.Sahoo, G.Arora, “Image thresholding using two-dimensional Tsallis–Havrda–Charvat entropy”, Pattern Recognition Letters, Vol. 27, pp. 520–528, 2006.

[19] J. Xue and D. M. Titterington (2011) “t-tests, F-tests and Otsu’s Methods for Image Thresholding,” IEEE Trans. Image Processing, vol. 20, no. 8, pp. 2392-2396.

[20] H. Tizhoosh, (2005) “Image thresholding using type II fuzzy sets”, Pattern Recognition, vol. 38, pp. 2363 – 2372.