Global Image Thresholding Based on Change-point Detection

Rohit Kamal Chatterjee, *Member, IAENG*, Avijit Kar

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

***Index Terms*— Change-point detection, Global image thresholding, Image segmentation, Kullback-Leibler divergence, Robust statistical measure**

I. Introduction

A gray-level digital image is a two dimensional signal, $\boldsymbol{I:\Omega\rightarrow L}$, where $\boldsymbol{\Omega}\subset\mathbb{Z}X\mathbb{Z}$, and $\boldsymbol{L}=\{l_i\in\mathbb{R}$ 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 $B=\{b_b(x,y)=1|I(x,y)<t_{0,}(x,y)\subset\boldsymbol{\Omega}\}$ and foreground $F=\{b_f(x,y)=1|\ I(x,y)\geq\ t_0,\ (x,y)\subset\boldsymbol{\Omega}\ \}$ or the opposite. The function $I(x,y)$ can take any random value $l_i\in\boldsymbol{L}$; 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 problem guided by the image histogram. In most of the cases, global thresholding methods try to estimate the threshold (t_0) iteratively by optimizing a criterion function [1]. Yet 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] and [20]. Comprehensive surveys discussing various aspects of thresholding methods can be found in the references [1], [6], and [7].

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.

We propose in this paper a novel algorithm addressing these drawbacks. Our algorithm uses a well-known statistical technique called *change-point detection (CPD)*. For the last few decades researchers in statistics and control theory have been attracted by the problem of detecting abrupt changes in the statistical behaviour of an observed signal or time series [9]. They are collectively called statistical change-point detection.

The general technique 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. Basically, it is assumed in CPD that θ takes values θ_0 and θ_1 ($\neq\theta_0$) before and after the *unknown change time* t_0.The problem is then to estimate this change in the parameter and the change time t_0. A very important property of the log-likelihood ratio $\boldsymbol{\Lambda(y)=log((p_{\theta_1}(y))/(p_{\theta_0}(y)))}$, becomes a tool for reaching this goal. Let $\boldsymbol{E_{\theta_0}}(\Lambda)$ and $\boldsymbol{E_{\theta_1}}(\Lambda)$ denote the expectations of Λ with respect to two distributions $\boldsymbol{p_{\theta_0}}$ and $\boldsymbol{p_{\theta_1}}$; then it can be shown that $\boldsymbol{E_{\theta_0}}(\Lambda)<0$ and $\boldsymbol{E_{\theta_1}}(\Lambda)>0$. In other words, a change in the parameter θ is reflected as a change in the sign of the expectation of the log-likelihood ratio. This statistical property can be used to detect the change in θ [10]. Given the *Kullback-Leibler (KL)* divergence $\boldsymbol{K(p_{\theta_1}\|p_{\theta_0})=E_{\theta_1}}(\Lambda)$, difference between the two mean values is

$$\boldsymbol{E_{\theta_1}(\Lambda)-E_{\theta_0}(\Lambda)=K(p_{\theta_1}\parallel p_{\theta_0})+K(p_{\theta_0}\parallel p_{\theta_1})} \qquad (1)$$

Manuscript received December 23, 2016; revised January 29 2017.

Rohit Kamal Chatterjee is with the Department of Computer Science and Engineering of Birla Institute of Technology, Mesra, India (mobile: +91-9433-519-475; e-mail: rkchatterjee@ bitmesra.ac.in).

Avijit Kar is with the Department of Computer Science and Engineering of Jadavpur University, Kolkata, India. (e-mail: avijit_kar@cse.jdvu.ac.in).

From this we infer that the detection of a change can also be made with the help of the KL divergence before and after the change. This concept is used in this paper for deciding the threshold in an image histogram.

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 global thresholding as a change-point detection problem, section 4 presents the experimental results and compares the results with various often cited global thresholding algorithms; and finally section 5 summarizes main ideas in this paper.

II. The Change-Point Detection (CPD) Problem

Change-point detection (CPD) problem can be classified into two broad categories: real-time or online change-point detection, which targets applications where 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]. Since the image and the corresponding histogram are available to use, we concentrate on offline change-point detection in this paper. We also assume that there is only one change point throughout the given observations $\{y_k\}_{1\dots n}$. 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.

A. Problem Statement

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

$H_0: \varphi_j = \theta_0, for\ 1 \le j \le n$ versus an alternative hypothesis

$$H_1: \varphi_j = \begin{cases} \theta_0, for\ 1 \le j \le k-1 \\ \theta_1, for\ k \le j \le n. \end{cases} \tag{2}$$

where $\theta_0 \neq \theta_1$ and change time k is not known. 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

$$\Lambda_1^n(k) = \frac{\prod_{j=1}^{k-1} p_\theta(y_j) \times \prod_{j=k}^{n} p_{\theta'}(y_j)}{\prod_{j=1}^{n} p_{\tilde{\theta}}(y_j)} \tag{3}$$

where p_θ and $p_{\theta'}$ are pdfs before and after the change occurs and $p_{\tilde{\theta}}$ is the overall probability density. When the only unknown parameter is t_0, its maximum likelihood estimate (MLE) is given by the following statistic

$$\hat{t}_0 = \arg\max_{1\le k\le n} \Lambda_1^n(k) \tag{4}$$

B. Offline estimation of the change time

Considering equation (4) and (5) and the fact that $\prod_1^n p_{\tilde{\theta}}$ is a constant for given data, the corresponding MLE estimate is

$$\hat{t}_0 = \arg\max_{1\le k\le n} \ln\left(\prod_{j=1}^{k-1} p_\theta(y_j) \times \prod_{j=k}^{n} p_{\theta'}(y_j)\right) \tag{5}$$

where $\hat{t}_0$ is maximum log-likelihood estimate of t_0. Rewriting equation (6) as

$$\hat{t}_0 = \arg\max_{1\le k\le n} \left(\ln\left(\frac{\prod_{j=k}^{n} p_{\theta'}(y_j)}{\prod_{j=1}^{n} p_\theta(y_j)}\right) + \ln\left(\prod_{j=1}^{n} p_\theta(y_j)\right)\right) \tag{6}$$

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

$$\hat{t}_0 = \arg\max_{1\le k\le n} \sum_{j=k}^{n} \ln\left(\frac{p_{\theta'}(y_j)}{p_\theta(y_j)}\right) \tag{7}$$

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 value given by equation (7).

III. Global Thresholding: A Change-Point Detection Formulation

A. Assumptions

If Let $(\chi, \beta_\chi, P_\theta)_{\theta\in\Theta}$ be the *statistical space of discrete grey-levels* associated with a random variable $Y{:}\,\mathbb{Z}x\mathbb{Z}\to\mathbb{Z}$, where β_χ is the σ-field of Borel subsets $A \subset \chi$ 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 of all grey-level images with n elements that could be classified into M categories or classes $L=\{l_1, \dots, l_M\}$, i.e. each sample point in the sample image can take any random grey-level values from the set L.

B. Change-point detection formulation

Since we are mainly interested in discrete grey-level data, we consider the *multinomial distribution* model. Let $\wp=\{E_i\}$, i=1,...,M be a partition of χ. The formula $Pr_\theta(E_i) = p_i(\theta)$, $i = 1, \dots, M$, defines a discrete statistical model with probability of the l_i^{th} grey-level. Further we assume $\{Y_1, \dots, Y_n\}$ to be a random sample from the population described by the random variable $\mathbf{Y}$, representing the grey-level of a pixel. And let $N_i = \sum_{j=1}^{n} I_{E_i}(Y_j)$, where I_E is the index function. Then we can approximate $p_i(\theta) \approx N_i/n$, $i=1,\dots,M$. Estimating θ by maximum likelihood method consists of maximizing the joint probability distribution for fixed $n_1, \dots, n_M$

$$\Pr_\theta(N_1 = n_1, \dots, N_M = n_M) = \frac{n!}{n_1! n_2! \dots n_M!}\left((p_1(\theta))^{n_1} (p_1(\theta))^{n_2} \dots (p_1(\theta))^{n_M}\right) \tag{8}$$

or equivalently maximizing the log-likelihood function

$$\Lambda(\theta) = \ln\left(\frac{n!}{n_1! n_2! \dots n_M!}\left((p_1(\theta))^{n_1} (p_1(\theta))^{n_2} \dots (p_1(\theta))^{n_M}\right)\right) \tag{9}$$

ISBN: 978-988-14047-3-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Therefore, referring to equation (5), problem of estimating the threshold by MLE can be stated as

$$\hat{t}_0 = \underset{1\le k\le M}{\arg\max} \ln\left(\frac{n!}{n_1!n_2!\ldots n_M!}\prod_{i=1}^{j-1}\left(p_i(\theta_0)\right)^{n_i}\times\prod_{i=j}^{M}\left(p_i(\theta_1)^{n_i}\right)\right) \quad (10)$$

where unknown parameter $\theta = \theta_0$ before change and $\theta = \theta_1$ after the change. Now, equation (10) can be expanded as

$$\hat{t}_0 = \underset{1\le j\le M}{\arg\max}\left[\ln\left(\frac{n!}{n_1!n_2!\ldots n_M!}\right)+\ln\left(\frac{\prod_{i=j}^{M}(p_i(\theta_1))^{n_i}}{\prod_{i=j}^{M}(p_i(\theta_0))^{n_i}}\right)+\ln\left(\prod_{i=1}^{M}(p_i(\theta_0))^{n_i}\right)\right] \quad \ldots (11)$$

The first term within the bracket on the right side equal sign of equation (11) is a constant and the last term is independent of j, i.e. it cannot influence the MLE. So, eliminating these terms and simplifying we get

$$\hat{t}_0 = \underset{1\le j\le M}{argmax}\sum_{i=j}^{M} ln\left(\frac{p(\theta_1)}{p(\theta_0)}\right)^{n_i} \quad (12)$$

Multiplying and dividing n on right side of equation (12)

$$\hat{t}_0 = \underset{1\le j\le M}{\arg\max}\, n\sum_{i=j}^{M}\left(\frac{n_i}{n}\right)\ln\left(\frac{p_i(\theta_1)}{p_i(\theta_0)}\right)^{n_i} \quad (13)$$

approximating $p_i(\theta)\approx n_i/n$ equation (13) can be written as

$$\hat{t}_0 \approx \underset{1\le j\le M}{\arg\max}\, n\sum_{i=j}^{M} p_i(\theta_1)\ln\left(\frac{p_i(\theta_1)}{p_i(\theta_0)}\right)^{n_i} \quad (14)$$

The expression in (14) under the summation denotes *Kullback-Leibler (KL)* divergence between the density $p(\theta_1)$ and $p(\theta_0)$; therefore equation (14) can be written as

$$\hat{t}_0 = \underset{1\le j\le M}{\arg\max}\, nK_j^M(p(\theta_1)\,||\,p(\theta_0)) \quad (15)$$

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 (15) can be rewritten as

$$\hat{t}_0 = \underset{1\le j\le M}{\arg\min}\, nK_1^j(p(\theta_1)\,||\,p(\theta_0)) \quad (16)$$

Hence, equation (16) provides the maximum likelihood estimation of the threshold t_0. Equation (16) can be restated as the follows: *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]

$$D(p_{\theta_1}\,//\,p_{\theta_0}) = \frac{1}{2}\left(K(p_{\theta_1}\,//\,p_{\theta_0}) + K(p_{\theta_0}\,//\,p_{\theta_1})\right) \quad (17)$$

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 generalised divergence measures in references [11] and [12].

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

TABLE I
COMPARISON OF RESULTS OF TEST IMAGES USING SSIM FOR THE IMAGES (A) DICE, (B) RICE, (C) OBJECT,

Image	Algorithm	Threshold	SSIM
Dice	Entropy	157	0.6705
	Kittler	55	0.9182
	Kurita	106	0.9560
	Otsu	103	0.9469
	Sahoo	134	0.7040
	CPD	96	0.9907
Rice	Entropy	118	0.9377
	Kittler	132	0.8172
	Kurita	126	0.8709
	Otsu	124	0.8880
	Sahoo	133	0.8082
	CPD	107	0.9748
Object	Entropy	150	0.8523
	Kittler	83	0.5381
	Kurita	107	0.7752
	Otsu	110	0.7638
	Sahoo	154	0.9014
	CPD	190	0.9014

IV. EXPERIMENTAL RESULTS AND DISCUSSION

To validate the applicability of proposed Change-Point Detection (CPD) thresholding algorithm, we provide experimental results and compare the results with existing algorithms. First row of **Fig.1** shows test images that are labelled from left to right as: *Dice, Rice, Object* respectively and the second row shows corresponding ground truth images. We have also selected few more test images shown in first row of **Fig.2** with names *Denise*, *Train*, and *Lena* to visually compare the results. These images have deliberately been so selected that the difference of areas between foreground and background are 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].

For evaluating the proposed algorithm, Structural Similarity Index (SSIM) [21] of the thresholded test images with respect to the ground truth images has been computed. **Table-1** shows optimal thresholds of five selected algorithms and the CPD algorithm. It is clear that CPD performs reasonably well.

For visual comparison, consider the Denise and Train image, Kittler-Illingworth thresholding totally fails to distinguish the object from background due to its assumption of Gaussian distribution for both foreground and background [19], similar consequence occurs with other four algorithms. But proposed CPD algorithm segments the images fairly well.

V. Conclusion

In this paper we propose a novel global image thresholding algorithm based on Statistical Change-Point detection (CPD), the derivation uses 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.

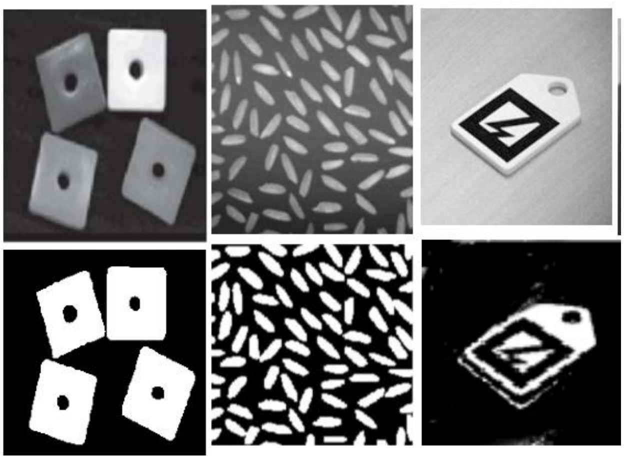

Fig 1: Four tested images: **Row-1**: Original images, Dice, Rice, and Object. **Row-2:** Ground truth images.

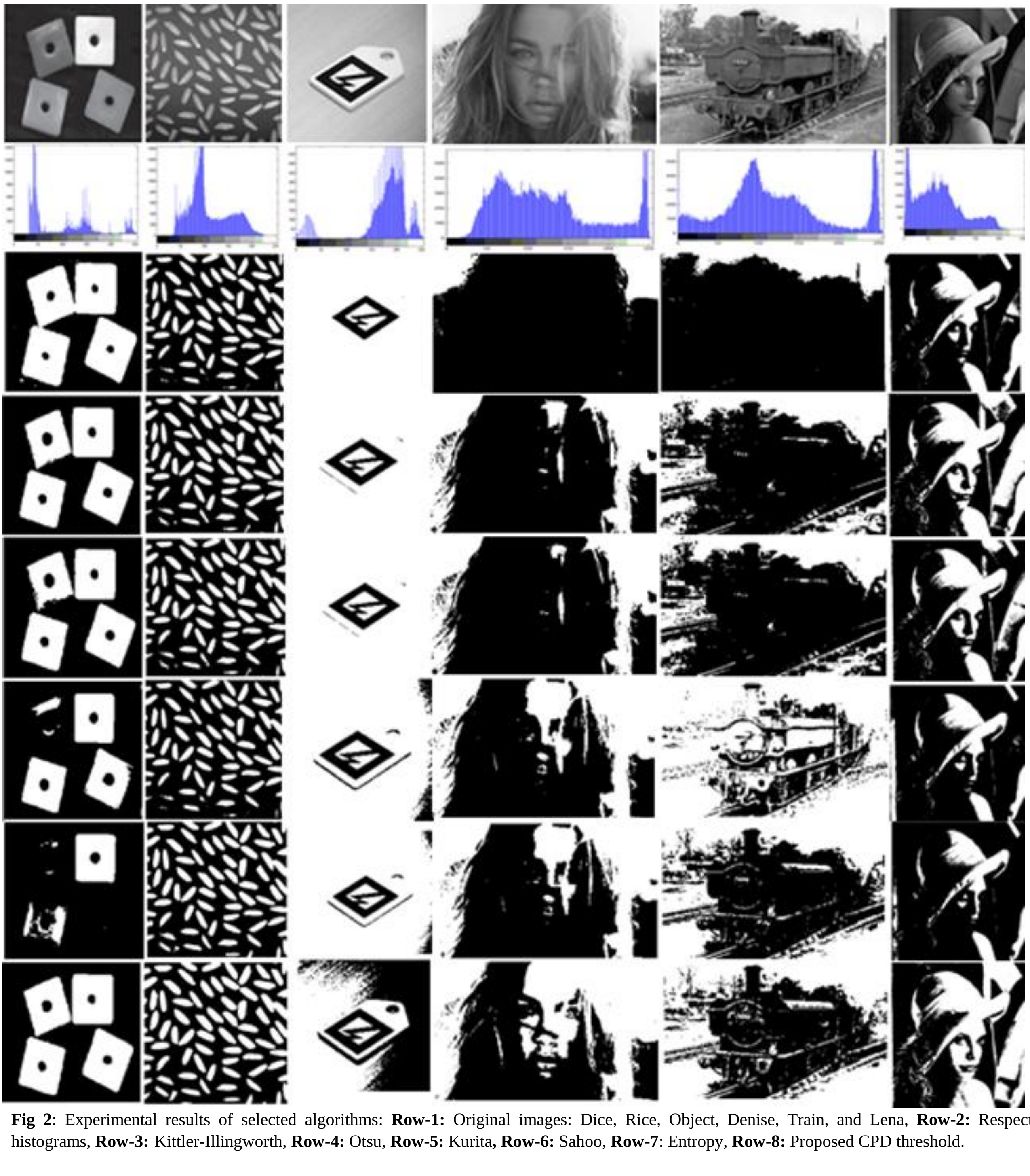

Fig 2: Experimental results of selected algorithms: **Row-1:** Original images: Dice, Rice, Object, Denise, Train, and Lena, **Row-2:** Respective histograms, **Row-3:** Kittler-Illingworth, **Row-4:** Otsu, **Row-5:** Kurita**, Row-6:** Sahoo, **Row-7**: Entropy, **Row-8:** Proposed CPD threshold.

ISBN: 978-988-14047-3-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

REFERENCES

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

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

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

[4] D. M. Tsai, ''A fast thresholding selection procedure for multimodal and unimodal histograms,'' *Pattern Recogn. Lett. vol.* 16, pp. 653–666, 1995.

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

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

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

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

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

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

[11] Y. Wang, "Generalized Information Theory: A Review and Outlook," *J. of Inform. Tech.*, vol. 10, no. 3, pp. 461-469, 2011.

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

[13] S. Choi, S. Cha, C. C. Tappert, "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, ''Minimum error thresholding,'' *Pattern Recogn.* Vol. 19, pp. 41–47, 1986.

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

[16] T. Kurita, N. Otsu, and N. Abdelmalek, "Maximum likelihood thresholding based on population mixture models". *Pattern Recogn.*, vol. 25, pp. 1231-1240, 1992.

[17] P. Sahoo, C. Wilkins, and J. Yeager, ''Threshold selection using Renyi's entropy,'' *Pattern Recogn.*, vol. 30, pp. 71–84, 1997.

[18] P.K.Sahoo, G. Arora, "Image thresholding using two-dimensional Tsallis–Havrda–Charvat entropy," *Pattern Recog. Lett.* Vol. 27, pp. 520–528, 2006.

[19] J. Xue and D. M. Titterington "t-tests, F-tests and Otsu's Methods for Image Thresholding," *IEEE Trans. Image Process.*, vol. 20, no.8, pp. 2392-2396, 2011.

[20] H. Tizhoosh, "Image thresholding using type II fuzzy sets," *Pattern Recog.*, vol. 38, pp. 2363 – 2372, 2005.

[21] Z. Wang et al, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions Image Processing.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

ISBN: 978-988-14047-3-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)