Comparison of Five Conditional Probabilities in 2-level Image Thresholding Based on Baysian Formulation

Yan Chang, Alan M. N. Fu, Hong Yan and Mansuo Zhao
School of Electrical and Information Engineering
The University of Sydney, NSW 2006, Australia
Phone +61-2-93516659, Fax: +61-2-9351-3847
E-mail: ychang@ee.usyd.edu.au

Abstract

An efficient method for two-level thresholding is proposed based on the Bayes' formula and the maximum entropy principle, in which no assumptions of the image histogram are made. Five forms of conditional probability distributions, *Simple*, *Linear*, *Parabola Concave*, *Parabola Convex* and *S-Function*, are employed and compared to each other for optimal threshold determination. The experiment results show that the *Parabola Concave* form is the most effective, retaining most of the information for most thresholding images. The *Linear* form is an acceptable form due to its simple one-order linear function.

1. INTRODUCTION

Image thresholding is an important and commonly used procedure in image segmentation. Many thresholding methods have been developed in the last two decades. Among them, the histogram-based thresholding technique [1-7] has been dominant since it only needs the gray-level histogram without other priori knowledge. While global thresholding methods, in which the entire image is thresholded with a single threshold value, have been widely applied because they are independent of the image size and also effective.

Based on Bayes' formula of probability theory, Ridler [2] and Lioyd [3] proposed the use of the thresholding algorithm which utilizes the criterion of maximizing the likelihood of the histogram. Mardia [4] improved this algorithm by taking into account the spatial relationship between neighboring pixels. In these algorithms, a Gaussian distribution is assumed within each partition area. The methods used by Ostu [5] and Kittler [6] are also based on the assumption of Gaussian distribution. Pun [7,8] developed the entropy-based method in the beginning of eighties, in which the most information is retained after an image has been thresholded . Yan [13] gives a unified formulation for the methods of Ostu [5], Kittle and Illingworth [6], and Huang and Wang [12]. The recently developed fuzzy theory is typically employed to select an optimal threshold by maximizing the fuzzy entropy [9-12]. In this paper, an efficient optimal thresholding method is proposed based on Bayes' formula and the maximum entropy principle. In this method, the optimal thresholds that preserve the most information of the partitioned image are affected by the form of conditional probability function of Bayes' formula. The *Linear* [10] and *S-Function* [11] forms are usually considered in fuzzy method. In this paper, we propose other two nonlinear functions, *Parabola Concave* and *Parabola Convex*. Unlike the likelihood method where the Gaussian distribution is assumed, no assumptions are made by our method.

The rest of this paper is organized as follows. In Section 2, the concept of the proposed method is described. In Section 2, we give the definitions of five forms of conditional probability distribution functions used to construct the entropy function *E* for selecting the optimal threshold value. Experiments are presented to verify the proposed method and compare five forms in Section 3. Finally, the conclusions are given in Section 4.

2. METHOD BASED ON BAYES' FORMULA AND MAXIMUM ENTROPY

Let $D \subset R^2$ denote the domain of an image and *G* denote the *L* gray levels of an image. Then, an image with *L* gray-levels is presented as

$$I = I(i, j) \in G \quad \text{for } (i, j) \in D. \qquad (1)$$

The histogram, denoted by $H = \{h_0, h_1, \dots h_{L-1}\}$, of an image *I* presents the frequency of occurrence of each gray level in the image and is obtained directly from the observation of the considered image.

Let $D_g = \{(i, j) \mid I(i, j) = h_g,\ (i, j) \in D\}$, $g \in G$ and n_g denote the total number of pixels in D_g. One can consider that each pixel $(i, j) \in D$ is selected with the probability $\frac{1}{N}$ where $N = N_x * N_y$ and it may belong to one of *L* classes D_g, $g \in G$. In view of this consideration, an image represents the results of a random experiment in which D_g denotes a random event with the probability $p_g \equiv P(D_g) = \frac{n_g}{N}$. Since $D = \bigcup_{g=0}^{L-1} D_g$ and $D_i \cap D_j = \Phi$, if $i \neq j$, where Φ denotes an empty set, hence $\Re = \{D_0, D_1, ..., D_{L-1}\}$ is a probability partition of *D*. Based on the definition of p_g, we have

$$p_g = h_g, \quad g = 0,1,...L-1. \tag{2}$$

Thus, the histogram *H* of an image determines the probability partition $\Re$ of *D*. Equation (2) presents the relationship between the histogram *H* and the probability partition $\Re$.

The aim of two-level thresholding of an image *I* is to separate its domain *D* into two parts, D_d^* and D_b^*, where D_d^* is composed of '*dark*' pixels corresponding to the smaller gray-level $g \in G$, and D_b^* is composed of those '*bright*' pixels corresponding to the larger gray-level $g \in G$. If the classification is achieved, then

$$\left. \begin{array}{l} D = D_d^* \cup D_b^* \\ D_d^* \cap D_b^* = \text{f} \end{array} \right\}. \tag{3}$$

This classification involves a determination of the optimal threshold $\tilde{g}$ on *G* such that a pixel $I(i, j)$ is classified into D_d^*, if $I(i, j) \leq \tilde{g}$; or D_b^*, if $I(i, j) > \tilde{g}$ under the condition that the information given by the original image is preserved as much as possible after this partition. However, due to the fact that the boundary between bright and dark is not well defined, some of the pixels with the same level (i.e. corresponding to the same $g \in G$) may be classified into D_d^* and others may be classified into D_b^*. This situation must be taken into account in determining the threshold value $\tilde{g}$. It is assumed therefore that for each $g \in G$, D_g is composed of two parts, D_{dg} and D_{bg}, where $D_{dg} \subset D_d^*$ and $D_{bg} \subset D_b^*$. Using the probability partitions $\Re$ and $\Re_g, g \in G$, we have

$$\left. \begin{array}{l} D_d^* = \bigcup_{g=0}^{L-1} D_{dg} \\ D_b^* = \bigcup_{g=0}^{L-1} D_{bg} \end{array} \right\}. \tag{4}$$

Let $P_d^* = P(D_d^*)$ and $P_b^* = P(D_b^*)$. Based on the complete probability formula, we therefore have

$$P_d^* = \sum_{g=0}^{L-1} p_g \cdot p_{d|g} \quad \text{and} \quad P_b^* = \sum_{g=0}^{L-1} p_g \cdot p_{b|g} \tag{5}$$

where

$$\left. \begin{array}{l} p_{d|g} = \dfrac{P(D_{dg})}{P(D_g)} \\ p_{b|g} = \dfrac{P(D_{bg})}{P(D_g)} \end{array} \right\}. \tag{6}$$

Note that $p_{b|g} = 1 - p_{d|g}$, therefore

$$P_b^* = 1 - P_d^* \tag{7}$$

In Equation (6), $p_{d|g}$ is the conditional probability of the event and $p_{b|g}$ is the conditional probability of the event D_b^*. From Equation (4) and (5), we can see that the two-level thresholding is determined by *L* sub-partitions $\Re_g, g \in G$. Based on Equation (6), the *L* sub-partitions are given by conditional probabilities $p_{d|g}$ and $p_{b|g}$, $g \in G$. Thus, based on the two relationships, it is clear that the two-level thresholding problem (i.e. to find partition $\{D_d^*, D_b^*\}$) is reduced to finding suitable conditional probability functions $p_{d|g}$ and $p_{b|g}$ with respect to the variable $g \in G$.

Based on the meanings of '*darkness*' and '*brightness*', as well as the relationship between them and gray-level measures, the functions $p_{d|g}$ and $p_{b|g}$ have the following properties:

$$\left. \begin{array}{l} 1 = p_{d|0} \geq p_{d|1} \geq ... \geq p_{d|(L-2)} \geq p_{d|(L-1)} = 0 \\ 0 = p_{b|0} \leq p_{b|1} \leq ... \leq p_{b|(L-2)} \leq p_{b|(L-1)} = 1 \end{array} \right\} \tag{8}$$

i.e. $p_{d|g}$ monotonically decreases and $p_{b|g}$ monotonically increases on *G*. The lower/higher the gray-level of the pixel is , the higher/lower the probability that the gray level belongs to darkness/ brightness.

We adopt the Shannon maximum entropy theory as the criterion of our method. Let $\Re_2^* = \{D_d^*, D_b^*\}$ be a probability partition of D, the entropy function E of partition $\Re_2^*$ is defined as

$$E = -P_d^* \lg P_d^* - P_b^* \lg P_b^* \qquad (9)$$

where

$$P_d^* = P(D_d^*) = \sum_{g=0}^{L-1} P(D_g) P(D_d^* \mid D_g) = \sum_{g=0}^{L-1} h_g\, p_{d|g}$$

$$P_b^* = P(D_b^*) = \sum_{g=0}^{L-1} P(D_g) P(D_b^* \mid D_g) = \sum_{g=0}^{L-1} h_g\, p_{b|g}$$

.

and $p_{d|g}$ and $p_{b|g}$ are given in Equation (6). Using Equation (7), Equation (9) is rewritten as

$$E = -P_d^* \lg P_d^* - (1 - P_d^*) \lg(1 - P_d^*) \,. \qquad (10)$$

Let $\tilde{p}_d$ is the optimal value that makes entropy E in equation (10) maximum. We can find

$$\tilde{p}_d = p_d^* = \frac{1}{2} \qquad (11)$$

3. Forms of $p_{d|g}$ and $p_{b|g}$

To determine the entropy in the space $\Re_2 = \{D_d, D_b\}$, the conditional probability functions $p_{d|g}$ and $p_{b|g}$ should have the property in equation (9). Following function form satisfies this property ($p_{b|g} = 1 - p_{d|g}$):

$$p_{d|g}(g,a,c) = \begin{cases} 1 & 0 \le g \le a \\ f(g,a,c) & a \le g \le c \\ 0 & c \le g \le L-1 \end{cases} \qquad (12)$$

We select five functions with 0-order, one-order and two-order respectively as follows (see Figure 1):

(a) *Simple*

$$p_{d|g}(g,a) = \begin{cases} 1 & 0 \le g \le a \\ 0 & a \le g \le L-1 \end{cases}$$

with one parameter *a*. This is actually a special case of the following four forms when *a* = *c*.

(b) Linear

$$p_{d|g}(g,a,c) = \begin{cases} 1 & 0 \le g \le a \\ \dfrac{g-c}{a-c} & a \le g \le c \\ 0 & c \le g \le L-1 \end{cases}$$

with two parameters *a* and *c*.

(c) *Parabola Convex* (with the apex in *a*):

$$p_{d|g}(g,a,c) = \begin{cases} 1 & 0 \le g \le a \\ \dfrac{-g^2 + 2ag + c(c-2a)}{(a-c)^2} & a \le g \le c \\ 0 & c \le g \le L-1 \end{cases}$$

with two parameters *a* and *c*.

(d) *Parabola Concave* (with the apex in *c*):

$$p_{d|g}(g,a,c) = \begin{cases} 1 & 0 \le g \le a \\ \dfrac{(g-c)^2}{(a-c)^2} & a \le g \le c \\ 0 & c \le g \le L-1 \end{cases}$$

with two parameters *a* and *c*.

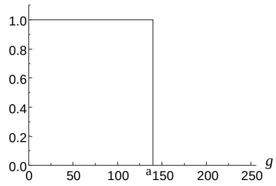
(e) *S-Function*

$$p_{d|g}(g,a,c) = \begin{cases} 1 & 0 \le g \le a \\ 1 - 2\left(\dfrac{g-a}{c-a}\right)^2 & a \le g \le \dfrac{a+c}{2} \\ 2\left(\dfrac{g-c}{c-a}\right)^2 & \dfrac{a+c}{2} \le g \le c \\ 0 & c \le g \le L-1. \end{cases}$$

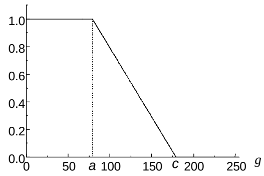
with two parameters *a* and *c*. This *S-Function* is symmetrical at the cross-over point $g = \frac{a+c}{2}$. The figure of above functions are shown in Figure 1.

4. EXPERIMENT RESULTS

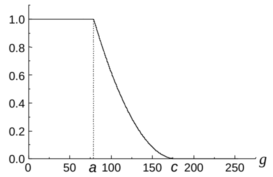
The experiments with bi-level thresholding on various kinds of images have been carried out with the proposed method using proposed five forms of conditional probability distribution functions. The four original images with various histogram distributions are selected and shown in Figure 2a-2d. Each image is presented by eight bits, that is, $L = 256$ from 0 (the darkest) to 255 (the brightest). Figure 3a-3d show the corresponding histograms of the images in Figure 2.



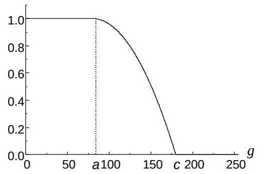
(a) *Simple*



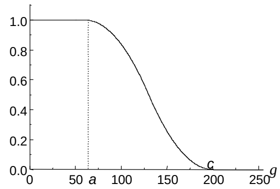
(b) *Linear*



(c)*Parabola Concave*



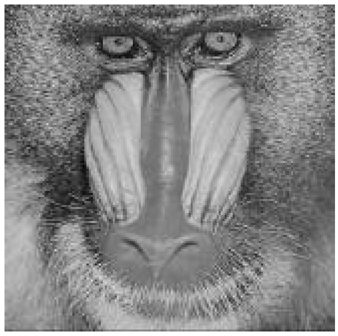
(d) Parabola Convex



(e) *S-Function*

Fig.1: Conditional probability function $p_{d|g}$

(a) *Airplane*



(b) *Baboon*

(c) *Lena*

(d) *Computer*

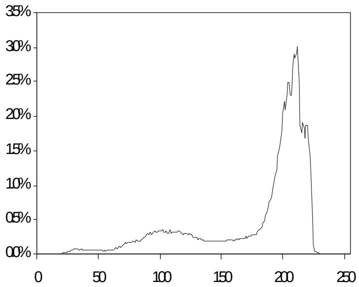
Fig.2 Orignal Images

The histograms of the images '*Airplane* (*512· 512*)' has only one peak in the brighter area and the distribution in the darker area is flat. Image '*Baboon (512· 512)*' is example of histogram with two clear peaks. The histograms of images '*Lena (512· 512)*' and '*Computer (640· 480)*' are of wide distribution and have more peaks. The contents of both images are difference types: human-face and in-door office scene.

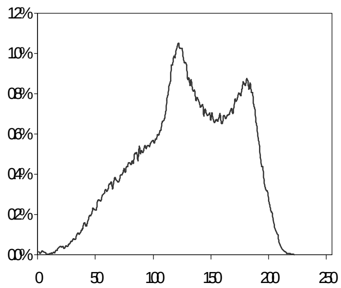
According to equation (11), we obtain the maximum entropy when $p_d^* = 0.5$. However, p_d^* is not equal to 0.5 exactly in most cases due to the discrete histogram. So we may use following minimum error *MinErr* instead of maximum entropy to find the optimal threshold $\tilde{g}$ in our experiment:

$$MinErr = \left| p_d - 0.5 \right| = \left| \sum_{g=0}^{c} p_{d|g} h_g - 0.5 \right|$$

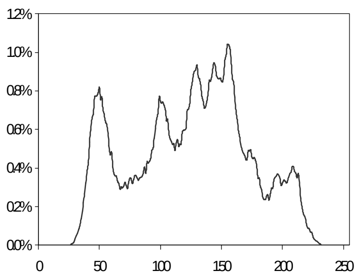
Table 1 gives the results of optimal thresholds $\tilde{g}$, corresponding parameters $(\tilde{a}, \tilde{c})$ and the minimum errors *MinErr* using the five forms of conditional probability functions $p_{d|g}$. It is obvious that the error from the *Simple* form is much higher than that from the other four forms for all images. However for some images, the *Simple* form obtains the same optimal



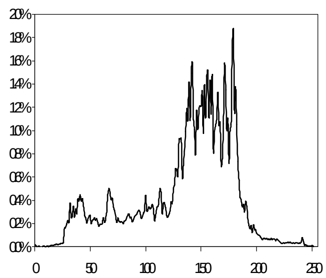
(a) *Airplane*



(b) *Baboon*



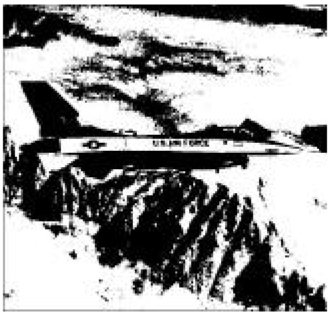
(c) Lena



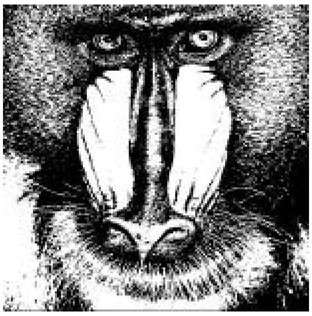
(d) Computer

Fig. 3. Histograms

threshold as some of the other forms. For example, the threshold *200* is obtained using both the *Simple* and *S-Funcion* forms for the *Airplane* image. Another example is for the *Submarine* image in which threshold *119* is obtained from the *Simple* and *Parabola Concave* forms. So the *Simple* form can not be ignored due to its brevity.

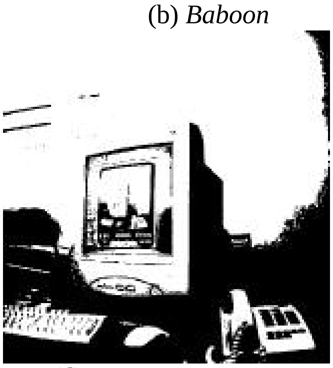


(a) *Airplane*



(b) *Baboon*

(c) *Lena*



(d) *Computer*

Figure 4: Thresholding Images (*Linear* form)

By comparing the five forms of $p_{d|g}$ from Table 1, it can be seen that the *Linear* and *Parabola* forms (*Concave* and *Convex*) achieve a smaller error than the *S-Function* form and an even smaller error than the *Simple* form. The minimum errors for the five forms for each image are marked in bold font in Table 1. That is, the *Linear* form gives the best results for the images *Airplane* and *Submarine*, *Parabola Concave* gives the best results for *Lena* and *Building*, and *Parabola Convex* for both the *Baboon* and *Computer* images. Therefore, it appears that the *S-Function* form is not a better choice of $p_{d|g}$.

4. CONCLUSIONS

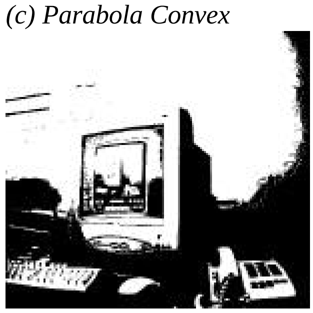
An efficient two-level thresholding method for monochrome images is proposed in this paper. The method is derived based on Bayes' formula and the maximum entropy principle in which the entropy function is given by a conditional probability function

(a) Simple

(b) Linear

(c) Parabola Convex

(d) Parabola Concave



(e) *S-Function*

Figure 5: Thresholding Image " *Computer*" with ε =0.001
(a) Simple form, $\tilde{g}$ *=146,* $(\tilde{a},\tilde{c})$ *=(146,146)*
(b) *Linear form,* $\tilde{g}$ *=135,* $(\tilde{a},\tilde{c})$ *=(30,240),*
(c) *Parabola Convex form,* $\tilde{g}$ *=142,* $(\tilde{a},\tilde{c})$ *=(91,244),*
(d) *(d) Parabola Concave form,* $\tilde{g}$ *=136,* $(\tilde{a},\tilde{c})$ *=(20,193),*
(e) *(e) S-Function form,* $\tilde{g}$ =138, $(\tilde{a},\tilde{c})$ *=(22,255).*

$p_{d|g}, g \in G$ and the histogram *H* of the considered image. The optimal threshold $\tilde{g}$ is obtained from the conditional probability function $\tilde{p}_{d|g}$ which makes the entropy function *E*($p_{d|g}$) arrive at the maximum value.

Table 1. Optimal thresholds and corresponding errors

Image		***Simple***	***Linear***	***Parabola Concave***	***Parabola convex***	***S-Function***
Airplane	$\tilde{g}$	200	183	196	186	200
	$(\tilde{a}, \tilde{c})$	(200, 200)	(118, 249)	(184, 225)	(25, 252)	(137, 247)
	MinErr(x10^{-5})	1269.53	**1.520**	13.82	1.926	1269.5
Baboon	$\tilde{g}$	130	131	130	133	131
	$(\tilde{a}, \tilde{c})$	(130, 130)	(84, 179)	(129, 132)	(91, 151)	(89, 174)
	MinErr(x10^{-5})	397.9	4.861	2.116	**0.016**	0.545
Lena	$\tilde{g}$	128	123	127	129	124
	$(\tilde{a}, \tilde{c})$	(128, 128)	(101, 153)	(121, 143)	(7, 180)	(9, 240)
	MinErr(x10^{-5})	190.7	1.996	**0.492**	1.654	0.645
Computer	$\tilde{g}$	146	135	136	142	138
	$(\tilde{a}, \tilde{c})$	(146, 146)	(30, 240)	(91, 244)	(20, 193)	(22, 255)
	MinErr(x10^{-5})	445.96	1.352	4.406	**0.624**	0.662

Compared with the maximum likelihood method [3-6], no assumption is made in our proposed method. To study the effect of the form of $p_{d|g}$, five forms of $p_{d|g}$ were used in the computation. The experiment results show that the *Parabola Concave* form is the most effective, retaining most of the information for most thresholding images. The *Linear* form is also an acceptable form due to its simple one-order linear function, while the *S-Function* form performs poorly and consists of two two-order nonlinear functions.

Reference

1. Y. Yang and H. Yan, An adaptive logical method for binarization of degraded document images, *Pattern Recognition*, 33(5): 787-807, 2000.
2. T.W.Riddler and S. Calvard, Picture thresholding using an iterative selection method, *IEEE Trans. Syst. Man Cybernet.* SMC-8, pp.630-632, Aug. 1978.
3. D.E. Lloyd, Automatic target classification using moment invariants of image shapes, Farnborough, UK, Rep. RAE IDN AW126. Dec. 1985.
4. K. V. Mardia and T. J. Hainsworth, A Spatial Thresholding Method for Image Segementation, *IEEE Trans. PAMI* Vol. 10, No. 6, NOV. pp.919-927 1988.
5. N.Otsu, A Threshold Selection Method from Gray-Level Histograms, *IEEE Trans. Syst. Man Cybernet.* SMC-9,62-66, 1979.
6. J. Kittler and J. Illingworth, Minimum Error Thresholding, *Pattern Recognition* Vol. 19 No.1 pp. 41-47, 1986.
7. T. Pun, A new method for gray level picture thresholding using the entropy of the histogram, *Signal Processing*, 2, pp. 223-237, 1980.
8. T. Pun, Entropic Thresholding, A New Approach, *Signal Processing*, 16, pp210-239, 1981.
9. T. G. Amaral, M. M. Crisostomo and A. T. de Almeida, Image Thresholding by Minimisation of Fuzzy Compactness and Linear Index of Fuzziness, *IEEE International Fuzzy System Conferece Proceedings*. August 22-25, Seoul Korea, 1999.
10. H. D. Cheng, J. R. Chen and J. Li, Threshold Selection Based on Fuzzy c-Patition Entropy Approach, *Pattern Recognition*, Vol. 31, No. 7 pp. 857-870, 1998.
11. H. D. Cheng, Yen-Hung Chen and Ying Sun, A novel fuzzy entropy approach to image enhancement and thresholding, *Signal Processing* 75, pp. 277-301, 1999.
12. L. K. Huang and M. J. Wang, Image thresholding by minimizing the measure of fuzziness, *Pattern Recognition* 28, 41-51 (1995).
13. H. Yan, Unified formulation of a class of optimal image thresholding techniques, *Pattern Recognition*, 29(12):2025-2032, 1996.