A Hierarchical Approach in Multilevel Thresholding Based on Maximum Entropy and Bayes' Formula

Yan Chang, Alan M. N. Fu and Hong Yan
School of Electrical and Information Engineering
The University of Sydney, NSW 2006, Australia

ychang@ee.usyd.edu.au

Abstract

An efficient hierarchical approach for image multi-level thresholding is proposed based on the maximum entropy principle and Bayes' formula, in which no assumptions of the image histogram are made. Five forms of conditional probability distributions are employed for optimal threshold determination. Our experiments demonstrate that the proposed method is effective and achieves a significant improvement in speed compared to the exhaustive search method.

1. Introduction

An important approach to image segmentation is gray level thresholding. It is a popular tool for computer vision and widely used in image analysis as a pre-processing tool. There are two kinds of thresholding methods: bilevel and mutilevel. For an image with clear objects in the background, the use of bilevel thesholding method is accepted. Over the years, many multilevel thresholding techniques have been developed. Some are direct extensions from bilevel thresholding techniques. But as the number of levels required increase, the complexity and computation time will also significantly increase.

Segmentation by peak detection methods is based on the clustering of gray levels around the peaks of the histogram to define homogeneous gray-level areas. In this approach, peaks and valleys are first detected from which thresholds are set to form gray-level clusters. Chang [1] uses a lowpass / highpass filter repeatedly to adjust (decrease/increase) the number of peaks or valleys to a desired number of classes and then the valleys in the filtered histogram are used as thresholds. Recently, fuzzy theory has been widely employed to select optimal thresholds by maximizing the fuzzy entropy [2-4]. However, the search schemes used by these methods, such as, exhaustive search and the genetic algorithm, are time consuming.

In this paper, an efficient hierarchical approach for multilevel thresholding is proposed based on the maximum entropy principle and Bayes' formula. Five forms of conditional probability functions with two parameters are determined automatically according to the histogram properties. The optimal thresholds obtained using the proposed method are the same as that using the exhaustive search method of the fuzzy entropy model, but our method is more efficient than existing entropy methods. Furthermore, unlike in the likelihood method where the Gaussian distribution is assumed, no assumptions are made by our method.

2. Principle based on Bayasian Formulation

Let D denote the two-dimensional intensity domain of an image I, and $G=\{0,1,\ldots L-1\}$ denote the L intensity values. Thus, an image I can be considered as a mapping from the two-dimensional domain D to the one-dimensional domain G, $I = I(i,j) \in G \quad for\ (i,j) \in D$

For an image I with L intensity values, we may consider it as L sub-spaces in G. $D_g = \{(i,j) \mid I(i,j) = g,\ (i,j) \in D\},\ g \in G$.

The purpose of multi-level thresholding of an image is to classify its L sub-spaces (intensity values) in G into K sub-spaces (intensity values). In general, $K<<L$. Let $D^* = \{D_k^* \mid k = 1,2,\ldots,K\}$ denote K sub-spaces. D_1^* is composed of the darkest pixels corresponding to the smaller intensity values and D_K^* is composed of those brightest pixels corresponding to the larger intensity values. If the classification is achieved, then

$$\left.\begin{array}{l} D^* = D_1^* \cup D_2^* \cup \ldots \cup D_K^* \\ D_1^* \cap D_2^* \cap \ldots \cap D_K^* = \phi \end{array}\right\} \qquad (1)$$

where ϕ denotes an empty set.

Due to the fact that the boundaries between the sub-spaces D_i^* and D_j^* ($i, j = 1,2,\ldots K$ and $i \neq j$) are not well defined, some of the pixels with the same sub-space (i.e. corresponding to the same $g \in G$) may be classified into the other. Therefore, it is assumed that for each $g \in G$, D_g is composed of K parts D_{kg}, $k=1,2,\ldots,K$, where $D_{kg} \in D_k^*$. $\Psi_g = \{D_{kg}, k = 1,2,\ldots,K\}$ is a probability partition of D_g and known as the *sub-partition*.

Using the probability partition Ψ_g, we have,

$$D_k^* = \bigcup_{g=0}^{L-1} D_{kg} \qquad k=1,2,\ldots,K.$$

Let $P_k^* = P(D_k^*)$, Based on the complete probability formula, we therefore have

$$P_k^* = \sum_{g=0}^{L-1} p_g \cdot p_{k|g} \ , \quad k=0,1,\ldots K. \qquad (2)$$

where

$$P_{k|g} = \frac{P(D_{kg})}{P(D_g)} \qquad (2a)$$

and

$$\sum_{k=0}^{K-1} p_{k|g} = 1 \qquad (3)$$

Instinctively, the closer the gray level g is to the k^{th} threshold t_i, the higher the conditional probability ($p_{k|g}$, $k = 0,1,2\ldots,K$) that g belongs to the k^{th} sub-space. The conditional probability $p_{k|g}$,therefore, has the following properties:

$$p_{k|t} \geq p_{k|t+1} \geq p_{k|t+2} \geq \ldots \geq p_{k|L-2} \geq p_{k|L-1}$$
$$p_{k|t} \geq p_{k|t-1} \geq p_{k|t-2} \geq \ldots \geq p_{k|1} \geq p_{k|0} \qquad (4)$$

3. Optimal thresholds

In partition space, the entropy is the sum of the entropies in K sub-spaces:

$$E(P_1^*, P_2^*, \ldots, P_K^*) = -P_1^* \lg P_1^* - P_2^* \lg P_2^* - \ldots - P_K^* \lg P_K^* \qquad (5)$$

where

$$\sum_{k=1}^{K} P_k^* = 1 \qquad (5a)$$

Thus the entropy function E is a functional of P_k^*, $k=1,2,\ldots K$, which has $K-1$ independent variables. Since P_k^* is defined by the conditional probability function $p_{k|g}$, $g \in G$ (see equation 2), entropy E is actually a functional of $p_{k|g}$ ($g \in G$ and $k=1,2,\ldots K-1$) and denoted by $E(p_{k|g} \mid k = 1,2,\ldots K)$. Therefore, the entropy $E(p_{k|g} \mid k = 1,2,\ldots K-1)$ can be considered as a measure of the compatibility between h_g and the conditional function of $p_{k|g}$, $g \in G$ and $k=1,2,\ldots K$. The larger the value of $E(p_{k|g} \mid k = 1,2,\ldots K-1)$, the more compatibility there is between h_g and $p_{k|g}$. Since $p_{k|g}$ should has the property shown in Equation (4), we assume that $p_{k|g}$ has the following form

For $1 < k \leq (K-1)/2$

$$p_{k|g}(g, a_k, c_k) = \begin{cases} 1 & 0 \leq g < a_k \\ f_k(g, a_k, c_k) & a_k \leq g < c_k \\ 0 & c_k \leq g < L \end{cases}$$

(6a)

and for $(K-1)/2 < k \leq K-1$

$$p_{k|g}(g, a_k, c_k) = \begin{cases} 0 & 0 \leq g < a_k \\ f_k(g, a_k, c_k) & a_k \leq g < c_k \\ 1 & c_k \leq g < L \end{cases}$$

(6b)

where $f_k(g, a_k, c_k)$ is a monotonous continuous function based on Equation (4). The optimal threshold in k^{th} sub-space $\tilde{t}_k$ is the one where $p_{k|g} = \frac{1}{2}$.

For obtaining the maximum value of Equation (5), we construct a new function:

$$\begin{aligned} E^*(P_1^*, P_2^*, \ldots, P_K^*) = \\ -P_1^* \lg P_1^* - P_2^* \lg P_2^* - \ldots \\ -P_K^* \lg P_K^* + \lambda(P_1^* + P_2^* + \ldots + P_K^* - 1) \end{aligned} \qquad (7)$$

where λ is the Lagrange constant.

Finding the derivative of Equation (6) to the variables $P_1^*, P_2^*, \ldots, P_{K-1}^*$, we have $K-1$ equations as follows

$$P_k^* = e^{(\lambda-1)}, \quad k = 1,2,\ldots,K-1 \qquad (8)$$

So P_k^* is a constant. Considering Equation (8) and (5a), we have

$$P_k^* = \frac{1}{K}, \quad k = 1,2,\ldots K \qquad (9)$$

Equation (9) shows the entropy E achieved at the maximum when the sub-spaces have the same probability $\frac{1}{K}$ which is the theoretical base of our paper.

4. Conditional Probabilities

As shown in Equation (6a) and (6b), the selected conditional probabilities $p_{k|g}$ with that a pixel locates in the k^{th} sub-space is with two parameters a_k and c_k . The total number of parameters for the entropy function are $2(K-1)$. All $p_{k|g}$, $k=1,2,\ldots,K$ are symmetrical in forms with the middle sub-space if K is odd, or the middle two sub-spaces if K is even. Many monotonous continuous functions could be selected as $f_k(g, a_k, c_k)$. The simplest of all possible functions is the one where $a_k = c_k$, that is, $f_k(g, a_k, c_k)$=0. In this case, the total number of parameters are decreased to $K-1$. For comparison, we name this form as the *simple* form in the experiment. We here select four other forms of $f_k(g, a_k, c_k)$ in the rang of $[a_k, c_k)$ as follows:

- *Linear*

$$f_k(g,a_k,c_k)=\begin{cases}\dfrac{g-c_k}{a_k-c_k} & 0<k\le(K\_1)/2\\ \dfrac{a_k-g}{a_k-c_k} & (K-1)/2<k<K\end{cases}$$

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

$$f_k(g,a_k,c_k)=\begin{cases}\dfrac{-g^2+2a_kg+c(c_k-2a_k)}{(a_k-c_k)^2} & 0<k\le(K-1)/2\\ \dfrac{(g-a_k)^2}{(a_k-c_k)^2} & (K-1)/2<k<K\end{cases}$$

- *Parabola Concave* (with the apex in *c*)

$$f_k(g,a_k,c_k)=\begin{cases}\dfrac{(g-c_k)^2}{(a_k-c_k)^2} & 0<k\le(K-1)/2\\ \dfrac{-g^2+2gc_k+a(a_k-c_k)}{(a_k-c_k)^2} & (K-1)/2<k<K\end{cases}$$

- *S-Function*

  When $0<k\le(K-1)/2$,

$$f_k(g,a_k,c_k)=\begin{cases}1-2\left(\dfrac{g-a_k}{c_k-a_k}\right)^2 & a_k<g\le\dfrac{a_k+c_k}{2}\\ 2\left(\dfrac{g-c_k}{c_k-a_k}\right)^2 & \dfrac{a_k+c_k}{2}<g\le c_k\end{cases}$$

  and when $(K-1)/2<k<K$,

$$f_k(g,a_k,c_k)=\begin{cases}2\left(\dfrac{g-a_k}{c_k-a_k}\right)^2 & a_k<g\le\dfrac{a_k+c_k}{2}\\ 1-2\left(\dfrac{g-c_k}{c_k-a_k}\right)^2 & \dfrac{a_k+c_k}{2}<g\le c_k\end{cases}$$

5. Hierarchical Approach

Assume $H_k=\{h_g^k \mid g\in D\}$ is the normalized histogram of the k^{th} sub-space. Based on the definition of p_g, Equation (2) can be rewritten as

$$P_k^*=\sum_{g=0}^{L-1}h_g^k\cdot p_{k|g}\ ,\qquad k=0,1,\ldots K. \qquad (10)$$

where $h_g^k=h_g^{k-1}p_{)k-1)|g}$ (11)

Thus, we develop a hierarchical approach based on these two equations. The D_g is first partitioned from darkness and brightness into the lowest and highest subspaces D_1^* and D_K^*, which is called the first layer. Then partition the second lowest and highest subspaces D_2^* and D_{K-1}^*, the second layer, and so on till the $\left(\dfrac{K}{2}\right)^{th}$ layer. Thus, each layer includes two gray levels except the last one. For the last layer, if K is odd, it only includes one gray level which is the middle subspace in D^*.

The procedures of our hierarchical approach are as follows:

Step1: computing the normalized histogram h_g^1 of the image

Step2: initializing the parameters of k^{th} sub-space by $a_k=0$ and $c_k=L-1$

1. computing $p_{k|g}$ by one of five forms in Section4.
2. computing P_k^* by equation (10).
3. if $\left|P_k^*-\dfrac{1}{K}\right|<\varepsilon$, where ε is a small number which indicates the precision, threshold $t_k=(a_k+c_k)/2$

   Otherwise, increasing a_k to $a_k=(a_k+c_k)/2$ if $(\dfrac{1}{K}-P_k^*)>\varepsilon$, or decreasing c_k to $c_k=(a_k+c_k)/2$ if $(P_k^*-\dfrac{1}{K})>\varepsilon$, then go to Step2.1 with a new (a_k,c_k)

Step3: modifying histogram h_g^k by Equation (11) and setting next sub-space *k=k+1*.

6. Experiment Results

We present the results of applying the proposed method to many kinds of images. Two of the original images and their histograms are presented in Figure 1. The sizes of the images *Ball* and *Mode*l are 392x432 and 277x574 respectively. Both are quantified to 256 gray levels (8bits).

Figure 2 shows the thresholded images with two-, three- and five-levels. We can see that the image qualities with multi-levels are much better than the one with two levels.

The results of five forms of conditional probabilities applied to the images *Ball* in three levels are shown in Figure 3. The five thresholds obtained by these five forms are quite different. The two thresholds in the *Linear* form (90,137) are quite different from the two thresholds in the *Simple* form (102, 124). This means the number of pixels in the middle sub-space by the *Linear* form is much bigger than that by the *Simple* form. Unlike two-level thresholding method used in [5], the five forms are not obviously different for higher level thresholding. Based on this experiment, when *K>5*, the objective results are quite similar.

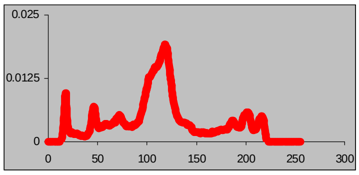
References

1. Cheng-Chia Chang and Ling-Ling Wang, A fast multilevel thresholding method based on lowpass and highpass filtering, Pattern Recognition Letters 18(1977) 1469-1478.
2. 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.
3. 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.
4. L. K. Huang and M. J. Wang, Image thresholding by minimizing the measure of fuzziness, *Pattern Recognition* 28, 41-51 (1995).
5. Yan Chang etc., Comparison of Five Conditional Probabilities in 2-level Image Thresholding Based on Baysian Formulation, Proceedings of VIP, 2000

Table 1. Thresholds and Parameters *a* and *c*

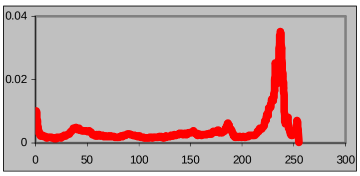
	Two-level	Three-level		Five-level			
Images	t (a,c)	t_1 (a_1,c_1)	t_2 (a_2,c_2)	t_1 (a_1,c_1)	t_2 (a_2,c_2)	t_3 (a_3,c_3)	t_4 (a_4,c_4)
Ball	106 (17,195)	90 (45,136)	137 (230,45)	77 (52,102)	98 (56,141)	118 (181,56)	157 (193,122)
Model	158 (61,255)	126 (50,203)	221 (246,197)	73 (13,134)	141 (33,249)	195 (249,141)	235 (250,220)

(a) Ball



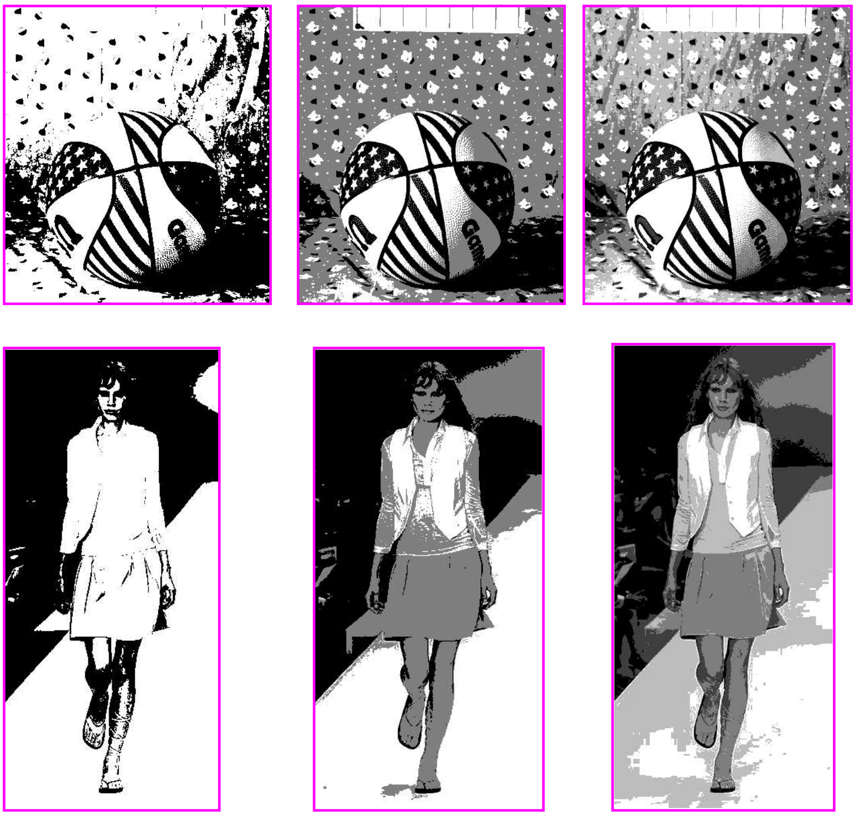
Histogram

(b) Model



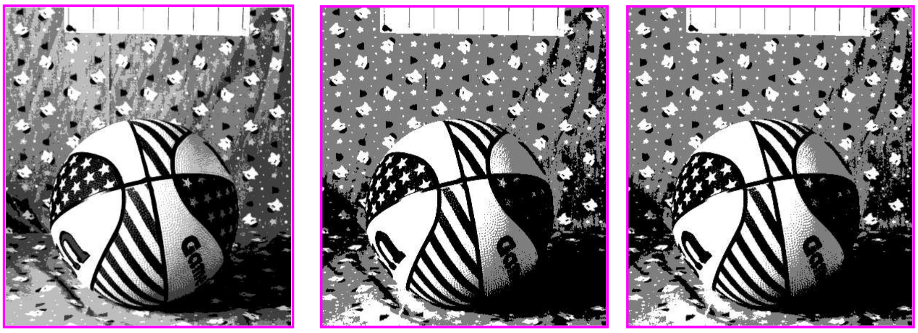
Histo gram

Figure 1. Original images

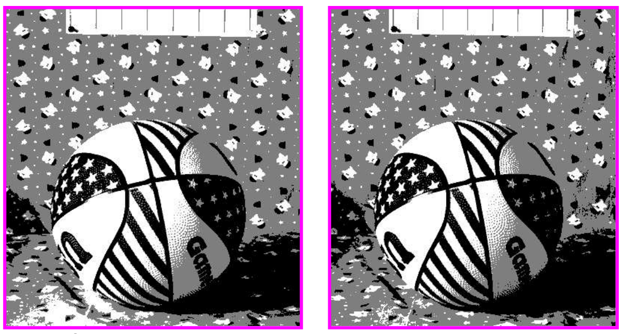


two levels three levels five levels

Figure 2 Thresholding images in Linear form



Linear (90,137) Concave (89,135) Convex (102,135)



S-function (95,144) Simple (102,124)

Figure3 Five forms of P_{ilg} at three-levels with two thresholds (t_1, t_2)