Automatic Threshold Selection Using Histogram Quantization

Yue Wang[1] *Tülay Adalı*[2] *and* *Shih-Chung B. Lo*[1]

[1]Center for Imaging Science and Information Systems
Georgetown University Medical Center, Washington, DC 20007
yuewang@isis.imac.georgetown.edu
[2]Department of Computer Science and Electrical Engineering
University of Maryland Baltimore County, Baltimore, MD 21228
adali@engr.umbc.edu

ABSTRACT

An automatic threshold selection method is proposed for biomedical image analysis based on a histogram coding scheme. We show that the threshold values can be determined based on the well-known Lloyd-Max scalar quantization rule, which is optimal in the sense of achieving minimum mean square error distortion. We derive an iterative self-organizing learning rule for determining the threshold levels which does not require any prior information about the histogram, and hence is fully automatic. Experimental results show that this new approach is very simple and efficient to implement yet yields reliable estimates of the threshold levels and is robust with respect to noise effect.

Keywords: Biomedical imaging, image analysis; multilevel thresholding; histogram quantization.

Correspondence: 2115 Wisconsin Avenue, NW, Suite 603, Washington, DC 20007.
Phone: (202) 687-0879, Fax: (202) 784-3479.

I Introduction

Thresholding is quite popular among a variety of image analysis techniques. This is primarily due to the fact that it is an easy yet efficient method to implement and provides satisfactory results in many cases. In various applications, it can be also effectively used as the initial step in more sophisticated image analysis tasks [1, 4]. Examples of such applications include segmentation of brain tissue and/or tumors in magnetic resonance (MR) images and quantification of nuclei of cells and chromosome in microscope images [8, 10]. However, poor contrast or strong noise component in the gray-level space of such images make thresholding a challenging task.

Thresholding assumes that the images present a number of relatively homogeneous regions, and that one can separate these regions by properly selecting the intensity thresholds [6]. Multi-level thresholding hence transforms the original image into a coarsely quantized image. Several threshold selection methods exist in the literature. These can be classified into two main classes: 1) histogram modeling and separation according to some specified criteria [3, 5] and 2) direct location of valleys and peaks in the histogram [2, 4, 6]. Histogram modeling often requires more sophisticated learning algorithms in order to obtain an unbiased estimate for the distribution model parameters [8]. On the other hand, in peak and valley detection, the sensitivity of the method to the noise level and the user defined control parameter becomes the main issue [6, 10]. Current approaches are all based on the *noisy* image histogram which is a sampled version of the true distribution, and employ a user-defined control parameter which allows the tuning of a series of trials to achieve the desired accuracy.

In this short report, we present an automatic threshold selection method based on a histogram coding scheme. We show that the threshold values can be determined based on the well-known Lloyd-Max scalar quantization rule, which is optimal in the sense of achieving minimum mean square error distortion. We derive an iterative self-organizing learning rule for determining the threshold levels which does not require any prior information about the histogram, and hence is fully automatic. Experimental results show that this new approach is very simple and efficient to implement compared to the similar approaches in [2, 5, 7] in terms of computational complexity (computational time and memory requirement), and yet yields reliable estimates of the threshold levels and is robust with respect to noise effect. Our recent study also shows the effectiveness of the proposed method in initializing a stochastic model-based image analysis algorithm in terms of resulting in faster rate of convergence and lower floor of local optimum likelihood.

II Self-Organizing Lloyd-Max Histogram Quantization

II.1 Problem Formulation

Suppose that an image is known to contain K regions and its pixels assume discrete gray level values u in the interval $[u_{MIN}, u_{MAX}]$. The distribution of the gray levels in the image can be approximated by a histogram $f(u)$ which gives the normalized frequency of occurrence of each gray level in the image. We formulate threshold selection as a histogram quantization problem which addresses the problem of determining the optimal coding scheme with $\log_2 K$ bits.

In rate distortion theory, Lloyd-Max scalar quantization has been proven to be optimal in the sense that it results in minimum distortion representation for a given distribution [10]. Following Max [9], we consider the histogram as a probability measure and define the global distortion measure D as the mean squared value of the quantization error. For a given number of regions, the coding scheme is described by specifying the thresholds t_k and the associated region means μ_k $(k = 1, ..., K)$ such that the global distortion D defined as

$$D = \sum_{k=1}^{K} \int_{t_k}^{t_{k+1}} (u - \mu_k)^2 f(u) du \tag{1}$$

is minimized. If we differentiate D with respect to t_k and μ_k and set the derivatives to zero, we get:

$$\frac{\partial D}{\partial t_k} = (t_k - \mu_{k-1})^2 f(t_k) - (t_k - \mu_k)^2 f(t_k) = 0, \quad k = 2, ..., K \tag{2}$$

and

$$\frac{\partial D}{\partial \mu_k} = 2 \int_{t_k}^{t_{k+1}} (u - \mu_k) f(u) du = 0, \quad k = 1, ..., K, \tag{3}$$

which yield

$$\mu_k = 2t_k - \mu_{k-1}, \quad k = 2, ..., K, \tag{4}$$

and

$$\int_{t_k}^{t_{k+1}} (u - \mu_k) f(u) du = 0, \quad k = 1, ..., K \tag{5}$$

where μ_k is the centroid of the area of $f(u)$ between t_k and t_{k+1}. This method provides a nice compromise between the profile and the details of the histogram, hence in general is not sensitive to noise effect. Note that since the method is applied to the original histogram that is actually a sampled version of a smooth probability density function, the thresholds and means do not necessarily correspond to the small valleys and peaks in the original histogram, and the goal is to find a noise-insensitive information representation such that a global distortion measure is minimized. Also, it is important to emphasize that it operates directly on the original histogram, i.e. no smoothing operation is needed which might lead to some loss of useful information.

II.2 Computation Algorithm

Because of the difficulty in obtaining an analytical closed-form solution for Eqs. (4) and (5), the problem can be attacked numerically. We propose the following procedure to calculate the mean and threshold levels in a complete unsupervised fashion: select an initial μ_1, calculate the corresponding t_ks and μ_ks for the K regions, and if μ_K is (or is close enough to) the true centroid of the last component μ_K^*, then μ_1 is chosen correctly; otherwise, update μ_1 as a function of the distance between μ_K and μ_K^*. For this update, we introduce a new parameter, α, to control the learning rate. The algorithm can be summarized as follows:

Self-Organizing Lloyd-Max Histogram Quantization (SLMHQ):

1. Initialization: Given K, set α, ϵ, and $m = 0$. Pick $\mu_1^{(0)}$.

2. For $k = 1, ..., K - 1$
   - set $t_1^{(m)} = u_{MIN}$
   - compute $t_{k+1}^{(m)}$ by

$$\sum_{u=t_k^{(m)}}^{t_{k+1}^{(m)}} u f(u) = \mu_k^{(m)} \sum_{u=t_k^{(m)}}^{t_{k+1}^{(m)}} f(u) \tag{6}$$

   - compute $\mu_{k+1}^{(m)}$ by

$$\mu_{k+1}^{(m)} = 2t_{k+1}^{(m)} - \mu_k^{(m)} \tag{7}$$

- set $t_{K+1}^{(m)} = u_{MAX}$ and compute $\mu_K^{*(m)}$ by Eq. (6)

3. If $|\mu_K^{(m)} - \mu_K^{*(m)}| < \epsilon$ then

   Go to step 4

   Otherwise

$$\mu_1^{(m+1)} = \mu_1^{(m)} + \alpha(\mu_K^{*(m)} - \mu_K^{(m)})$$

$$m = m + 1$$

   Go to step 2

4. Save the result and stop.

Note that after the initial guess, we compute the updates for μ_1 as a function of the distance between μ_K and μ_K^* computed in that iteration which results in a self organizing learning mechanism. The motivation leading to this rule can be briefly explained as follows. As a self-organizing approach, the correct selection of μ_1 depends on how close of μ_K to the true centroid μ_K^*. Thus, we can define $|\mu_K - \mu_K^*|$ as the error measure that is used as both feedback signal and stopping criterion in the learning rule. Specifically, when $\mu_K^* > \mu_K$ the value of μ_1 should be increased, otherwise (i.e., when $\mu_K^* < \mu_K$), the value of μ_1 should be decreased. In addition, according to adaptive control and processing theory, an appropriate learning rate is required. Step 3 in above learning rule embedded these considerations. The resulting algorithm thus provides an efficient and totally unsupervised threshold selection method, and since it minimizes a global distortion measure it also appears to be the most noise robust of the algorithms that we have studied. However, theoretical study on the convergence of the proposed algorithm has not been done, i.e., it has not been theoretically shown that the convergence of the learning rule is guaranteed. Instead, we have implemented a program that incorporated an empirically optimized learning rate and tree-structured error protection mechanism. Intensive experiments of the algorithm with various image characteristics has shown its effectiveness in practical applications that will be further explained in the following sections.

III Experimental Results

In this section, we present application of the new thresholding selection method to two real biomedical images from two different imaging modalities: the digital microscope image of a cell (Figure 1) and the magnetic resonance (MR) image of human brain tissue (Figure 2). The dynamic range of these images is 12 bits and their histograms are shown in Figures 3 and 4.

The choice of the learning rate indicates a tradeoff between convergence rate of the algorithm, and the residual error in the final parameter values. Also, it has to be chosen small enough to ensure stability. In our studies with 12 bit images, experimental results show that $\alpha = 1/K^3$ is a good value to achieve a suitable balance among these requirements.

To illustrate the general quantification scheme, we first consider the cell image, and its finite bit coded representation for $K = 3, 4, 5$. The SLMHQ algorithm presented in Section II.2 is implemented with $\alpha = 1/K^3$ and the accuracy threshold $\epsilon = \epsilon_0$. In our experiments, we set $\epsilon_0 = 0.5$. The corresponding results are plotted in Figures 5a-5c which show the original histogram together with the positions of the thresholds (short bins) and the corresponding means (high bins). It can be seen that with a fixed number of quantization levels, the locations of the thresholds are fairly accurate. The selectivity is increased as the number of levels (number of regions K) is increased. As specified by the underlying cell biology, for the cell image, the major components are nucleus, rough endoplasmic reticulum, smooth endoplasmic reticulum, and cell liquid; resulting in four final quantization levels. The segmented result directly using the threshold values for $K = 4$ is shown in Figure 6. When compared with the original image, it can be observed that this results in a quite plausible segmentation result. Also important to note is when the original histogram is noisy and the "peaks" of the histogram are difficult to identify, the proposed technique still yields quite satisfactory threshold determination. For this example, the second and third components are not represented by two distinguishable peaks in the histogram but they can be identified effectively (as shown in Figure 6) by the proposed SLMHQ scheme.

Table 1 provides a summary of the quantitative results for the microscope cell image quantification. For the thresholded image with four components, the threshold values, the component

Thresholds	Means	Variances
20	80	9.5
119	158	15.9
174	190	17.1
202	214	32.7
256		
MSE=75.18	CR = 4.5	

Table 1: The Summary of Histogram Quantization Results (Cell Image).

mean and variances are listed with the associated mean square error (MSE) distortion D and compression ratio (CR) values. The compression ratio is defined as

$$CR(f) = \frac{H(f_d)}{H(f)}. \tag{8}$$

where H denotes the entropy, f is the original histogram, and f_d is the quantized multinomial probability mass function.

We then employ SLMHQ to the MR brain image shown in Figure 2. Notice that the corresponding histogram for this image is very noisy and has a unimodal profile. The results given in Figure 7a-7c again show that the histogram quantization method is reliable, and capable of separating major regions without being influenced by noise. To establish clinical targeted analysis for the major brain tissue types, we use a brain tissue model, discussed in [10], to determine the number of major regions of interest in the image. In our case, we consider gray matter (GM), white matter (WM), cerebro-spinal fluid (CSF), and their pair-wise partial volume mixtures. Since partial volume pixels created by limited resolution are assumed not to be significant, we are only interested in the functional region partial volume mixtures that is essentially an anatomical feature of the brain tissues. It is evident from Figure 8 that the thresholded image, with $K = 5$, provides a quite satisfactory result in which major tissue types are well separated. Specifically, from dark to bright, they correspond to CSF, CSF/GM, GM, GM/WM, and WM.

It should be addressed that in both examples, biases and classification errors occurred because of the possible heavy overlaps among closer components. This means that a "shift" of mean values, a "shrink" of variance values, or a "noisy" segmentation of images will be shown in the final result of thresholding. This problem is an intrinsic defect of all thresholding methods when used for image quantification and segmentation. In our recent work [10], we have developed a framework by combining SLMHQ step with a stochastic model-based technique for image quantification and segmentation. Our experimental results show that the new threshold selection method can provide a good initialization for the follow-up stages, including the Expectation-Maximization (EM) algorithm for image quantification and the Contextual Bayesian Relaxation Labeling (CBRL) algorithm for image segmentation, such that the convergence rate is increased and the likelihood of being trapped in local optima is reduced.

Table 2 summarizes the comparative effects of the initializations by random selection and by SLMHQ scheme on the final quantification and segmentation of MR brain image. In quantification experiment, we use a standard finite normal mixture (SFNM) to model the true pixel density distribution and apply the EM algorithm to obtain the maximum likelihood estimate [8, 11]. The quantification error is measured by the global relative entropy (GRE) between the image histogram and the SFNM distribution. By setting a fixed lower bound of GRE value, we run 20 times of the EM algorithm with different random initializations, the mean value of the iterations required by EM is 67, while only 35 iterations, when using SLMHQ initialization, are sufficient for achieving the same accuracy. Furthermore, based on the initial thresholding result, we use CBRL algorithm to obtain final contextual segmentation [10]. Our test shows that, at the stationary point (no pixel re-labeling is required for the whole image), random initialization uses about 25 iterations of the CBRL, and SLMHQ initialization uses only 12 iterations of the CBRL. These results show that the SLMHQ initialization can increase the rate of convergence in both image quantification and segmentation. Secondly, we use the same set of random initializations and apply the EM algorithm with 1000 iterations. The results show that, in all cases, the EM algorithm reaches a stationary point with a GRE value of around 0.014 bits. On the other hand, when using SLMHQ initialization, the final GRE value is down to about 0.008 bits. This clearly provides us an evidence that SLMHQ initialization can reduce the likelihood of solution being trapped into local minima.

Furthermore, we also conducted a comparison study between the SLMHQ selection and Kohonen's self-organizing map (SOM) [7] and classification-maximization (CM) [8] algorithm since these two methods have also been frequently used for initializing image analysis algorithms

Items	Random Initialization	SLMHQ Initialization
Iterations of EM (GRE=0.087 bits)	67	35
Iterations of CBRL (Stationary Point)	25	12
Absolute GRE Values (1000 Iterations)	0.014 bits	0.008 bits

Table 2: The Comparison of Random and SLMHQ Initializations (MR Image).

Items	MSE	GRE (bits)
SLMHQ	75.18	0.039
SOM	86.29	0.143
CM	77.22	0.031

Table 3: The Comparison of SLMHQ/SOM/CM (Cell Image).

in many applications [7]. The evaluation criterion is a critical issue in our comparison since there is no gold standard. In this research, we used both the quantization error (MSE in the gray-level domain) and the quantification error (GRE in the probability domain) as the performance measures. The numerical results are given in Table 3. It can be seen that the SLMHQ outperform both SOM and CM algorithms in general (except the GRE value of CM result). The inferior performances of the SOM and CM algorithms may be explained as follows. In SOM, since the Euclidean distance is used for the competitive learning, only mean difference is taken into account such that the thresholds are the centroids of the means. This may be suitable only when the variances of all components are identical. CM algorithm uses a modified Mahalanobis distance to achieve a maximum likelihood classification which clearly improves the final results. However, as the prior probability of each component (e.g. prior in Bayesian classifier) is missing in the formulation, the method can not handle well the unbalanced mixture cases. In a contrast, the results of SLMHQ selection is most close to the Bayesian classification when the image histogram can be well modeled by a SFNM distribution [5, 10]. In addition, note that neither SOM nor CM is an automatic method since it still needs an initialization step which is eliminated in our approach.

IV Conclusion and Extended Work

In this paper, we present an automatic threshold selection method for image analysis and demonstrate the efficiency and reliability of the algorithm. The technique is unique in that it poses the problem as optimal scalar quantization of the image histogram, and seeks to minimize a global distortion measure to determine the optimum threshold levels. We have shown that the coarse-to-fine quantization of the information content of the histogram allows automatic selection of the number of threshold values to properly describe the dominant structures of the image at a given number of levels. The method is very promising in application to real medical images since 1) it is insensitive to the presence of noise in the histogram; 2) it can achieve a fully automatic search by using a self-organizing mechanism, no trial-and-error stage is required; and 3) it is an efficient computational procedure and hence can be implemented in real-time. We have extended our method to the initialization of hierarchical mixtures of experts neural network in computer-aided diagnosis where the feature space is two-dimensional. The preliminary results are very satisfactory.

Acknowledgment

The authors would like to thank Professor Tianhu Lei for his constructive comments which led to improve this work. The authors are also grateful to Ms. Susan Kirby for her editorial assistance.

References

[1] P. K. Sahoo, S. Soltani, A. K. C. Wong, and Y. C. Chen, "A survey of thresholding techniques," *Computer Vision Graphics Image Processing* **41**, 233-260 (1988).

[2] A. Rosenfeld and P. De La Torre, "Histogram concavity analysis as an aid in threshold selection," *IEEE Trans. Systems Man Cybernet.* **13**, 231-235 (1983).

[3] K. V. Mardia and T. J. Hainsworth, "A spatial thresholding method for image segmentation," *IEEE Trans. Pattern Analysis Machine Intell.* **10**(6), 919-927 (1988).

[4] M. I. Sezan, "A peak detection algorithm and its application to histogram-based image data reduction," *Computer Vision Graphics Image Processing* **49**, 36-51 (1990).

[5] T. Kurita, N. Otsu, and N. Abdelmalek, "Maximum likelihood thresholding based on population mixture models," *Pattern Recognition* **25**(10), 1231-1240 (1992).

[6] J. C. Olivo, "Automatic threshold selection using the wavelet transform," *Graphical Models and Image Processing* **56**(3), 205-218 (1994).

[7] J. L. Marroquin and F. Girosi, "Some extensions of the K-means algorithm for image segmentation and pattern classification," Technical Report, MIT Artificial Intelligence Laboratory, January 1993.

[8] Y. Wang and T. Lei, "A new stochastic model-based image segmentation technique for MR images," *Proc. 1st IEEE Intl. Conf. Image Processing*, Austin, Texas, 182-185 (1994).

[9] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory* **6**, 7-12 (1960).

[10] Y. Wang, T. Adali, M. T. Freedman, and S. K. Mun, "MR brain image analysis by distribution learning and relaxation labeling," (Invited Talk) *Proc. 15th Southern Biomed. Eng. Conf.*, March 29-31, Dayton, Ohio, pp. 133-136 (1996).

[11] Y. Wang, T. Adalı, and C. Lau, "Quantification of MR brain images by probabilistic self-organizing map," *Radiology*, Vol. 197(P), November (1995).