AUTOMATIC THRESHOLD SELECTION USING HISTOGRAM QUANTIZATION

Yue Wang,[†] Tülay Adali,[‡] and Shih-Chung B. Lo[†]
[†]Georgetown University Medical Center, Center for Imaging Science and Information Systems, Washington, DC 20007; [‡]University of Maryland Baltimore County, Department of Computer Science and Electrical Engineering, Baltimore, Maryland 21228

(Paper JBO-106 received July 30, 1996; revised manuscript received Jan. 21, 1997; accepted for publication Feb. 24, 1997.)

ABSTRACT

An automatic threshold selection method is proposed for biomedical image analysis based on a histogram coding scheme. 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. An iterative self-organizing learning rule is derived to determine the threshold levels. The rule does not require any prior information about the histogram, hence is fully automatic. Experimental results show that this new approach is easy to implement yet is highly efficient, robust with respect to noise, and yields reliable estimates of the threshold levels. *© 1997 Society of Photo-Optical Instrumentation Engineers.* [S1083-3668(97)00302-X]

Keywords biomedical imaging; image analysis; multilevel thresholding; histogram quantization.

1 INTRODUCTION

Thresholding is quite popular among a variety of image analysis techniques. This is primarily because it is easy yet efficient to implement and provides satisfactory results in many cases. In various applications, it can be also used as the initial step in more sophisticated image analysis tasks.[1,2] Examples of such applications include segmentation of brain tissue and/or tumors in magnetic resonance (MR) images and quantification of nuclei of cells and chromosomes in microscope images.[3,4] However, poor contrast or strong noise in the gray-level space of such images makes 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.[5] Multilevel thresholding hence transforms the original image into a coarsely quantized one. Several threshold selection methods are described in the literature. They can be grouped into two main classes: (1) histogram modeling and separation according to some specified criteria[6,7] and (2) direct location of valleys and peaks in the histogram.[2,5,8] Histogram modeling often requires more sophisticated learning algorithms to obtain an unbiased estimate for the distribution model parameters.[3] 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.[4,5] 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 that allows a series of trials to be tuned 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 (MSE) distortion. We derive an iterative self-organizing learning rule for determining the threshold levels that 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; i.e., it has low computational complexity (lower computational time and memory requirement) compared with similar approaches, such as those in Refs. 7 through 9, yields reliable estimates of the threshold levels; and is robust with respect to noise. Our recent study also shows the effectiveness of the proposed method in initializing a stochastic model-based image analysis algorithm in terms of leading to faster rate of convergence and lower floor of local optimum likelihood in the final quantification scheme.[4,10]

2 SELF-ORGANIZING LLOYD–MAX HISTOGRAM QUANTIZATION

2.1 PROBLEM FORMULATION

Suppose that an image is known to contain K regions and its pixels assume discrete gray-level val-

Address all correspondence to Yue Wang. E-mail: yuewang@isis.imac.georgetown.edu

1083-3668/97/$10.00 © 1997 SPIE

ues u in the interval $[u_{\mathrm{MIN}}, u_{\mathrm{MAX}}]$. The distribution of the gray levels in the image can be approximated by a histogram $f(u)$ that gives the normalized frequency of occurrence of each gray level in the image. We formulate threshold selection as a histogram quantization problem that 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 of representation for a given distribution.[4] Following Max,[11] 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 $(k=1,\ldots,K)$ so that the global distortion D defined as

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

is minimized. If we differentiate D with respect to t_k and μ_k and set the derivatives to 0, we get:

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

$$k=2,\ldots,K \tag{2}$$

and

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

which yields

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

and

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

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, which 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 so 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 involved, since the smoothing might lead to some loss of useful information.

2.2 Computation Algorithm

Because of the difficulty of 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 $\alpha$, $\epsilon$, and $m = 0$. Pick $\mu_1^{(0)}$.
2. For $k=1,\ldots,K-1$
   • Set $t_1^{(m)} = u_{\mathrm{MIN}}$.
   • Compute $t_{k+1}^{(m)}$ by
        $\sum_{u=t_k^{(m)}}^{t_{k+1}^{(m)}} uf(u)=\mu_k^{(m)}\sum_{u=t_k^{(m)}}^{t_{k+1}^{(m)}} f(u).$     (6)
   • Compute $\mu_{k+1}^{(m)}$ by
        $\mu_{k+1}^{(m)}=2t_{k+1}^{(m)}-\mu_k^{(m)}.$     (7)
   • Set $t_{K+1}^{(m)}=u_{\mathrm{MAX}}$ 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 the update rule in step 3 can briefly be explained as follows: As a self-organizing approach, the correct selection of μ_1 depends on how close μ_K is to the true centroid μ_K^*. Thus, we can define $|\mu_K-\mu_K^*|$ as the error measure that is used as both the feedback signal and the 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 the update, the positive constant α controls the amount of feedback; i.e., it determines the learning rate. The resulting algorithm thus provides an efficient and totally unsu-

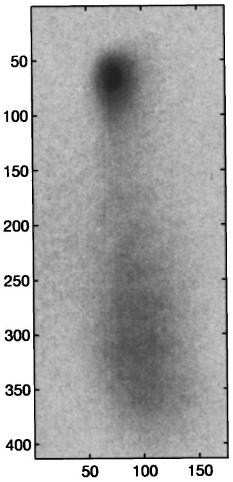


Fig. 1 Digitally imaged internal cell structures acquired by a CCD microscopic system.

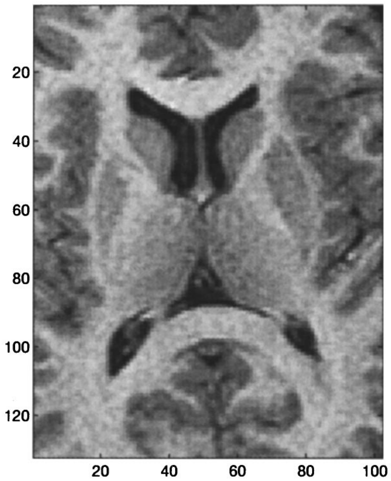


Fig. 2 A typical slice view of magnetic resonance brain tissue images.

pervised threshold selection method and since it minimizes a global distortion measure, it is also observed to be the most noise robust of the algorithms that we have studied. However, theoretical study of the convergence of the proposed algorithm has not been done; i.e., it has not been shown that the convergence of the learning rule is guaranteed. Instead, we have implemented a program that incorporates an empirically optimized learning rate and a tree-structured error protection mechanism.[10] Intensive numerical experiments with various image characteristics have shown the effectiveness of the algorithm in practical applications that we explain further in the following sections.

3 Experimental Results

In this section we present an 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 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 the 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, the 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, and 5. The SLMHQ algorithm presented in Sec. 2.2 is implemented with $\alpha=1/K^3$ and the stopping threshold is chosen as $\epsilon=0.5$. The corresponding results are plotted in Figures 5(a) through 5(c), 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 that directly uses the threshold values for $K = 4$ is shown in Figure 6. When compared with the original image, it can be observed that this achieves a quite plausible segmentation result. Also important to note is the point that when the original histogram is noisy and the "peaks" of the histogram are difficult to identify, the proposed technique still

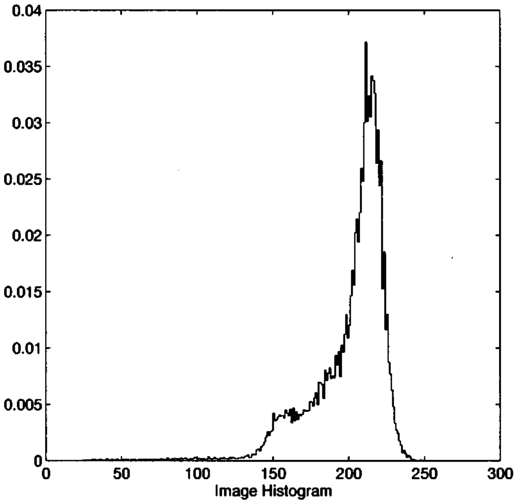


Fig. 3 Histogram of the cell image in Figure 1.

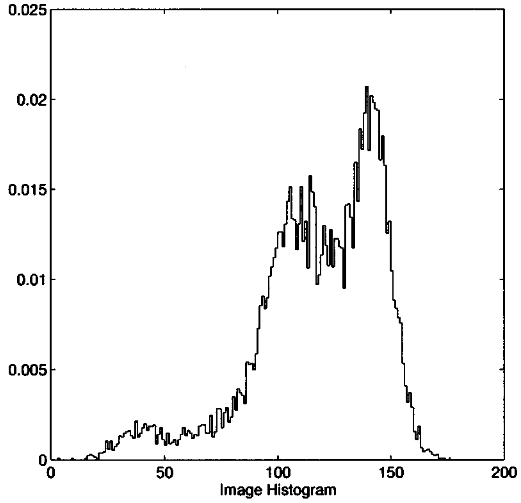


Fig. 4 Histogram of the MR brain image in Figure 2.

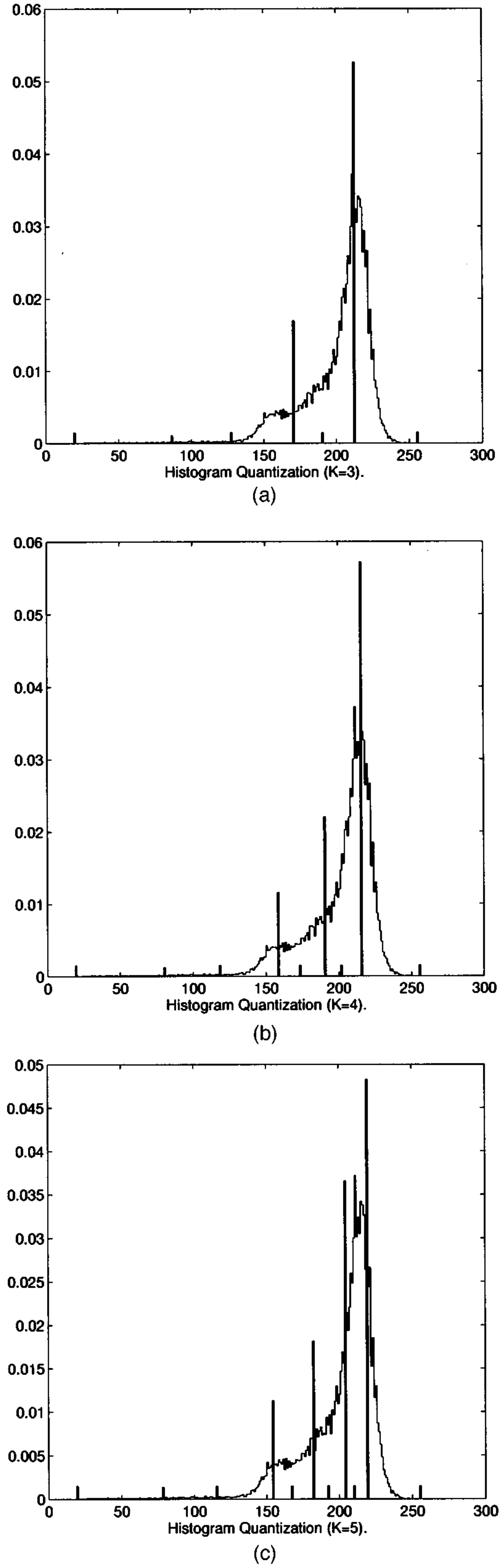
yields quite satisfactory threshold determination results. For this example, the second and third components are not observed as two distinguishable peaks in the histogram, but they can be identified effectively (as shown in Figure 6) by the proposed SLMHQ scheme.

Table 1 provides a summary of the quantitative results of the microscope cell image quantification. For the image thresholded with four components, the threshold values, the component mean, and variances are listed with the associated mean-square-error 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 apply 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 Figures 7(a) through 7(c) again show that the histogram quantization method is reliable and capable of separating major regions without being influenced by noise. To establish clinical targeted analyses for the major brain tissue types, we used a brain tissue model, discussed in Ref. 4, to determine the number of major regions of interest in the image. In our case, we considered gray matter (GM), white matter (WM), cerebrospinal fluid (CSF), and their pairwise partial volume mixtures. Since partial volume pixels created by limited resolution are assumed to be not significant, we are only interested in the functional region partial volume mixtures that are essentially an anatomical feature of the brain tissues. It is evident from Figure 8 that the thresholded image, with



(a)

(b)

(c)

Fig. 5 Results of the optimal quantization of a given histogram (cell), where (a), (b), and (c) correspond to the three-, four-, and five-level quantization, respectively. Note that the higher bins in the figures represent the centroids and the lower bins represent the thresholds.

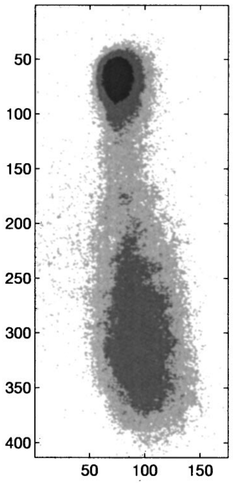


Fig. 6 Result of a simple thresholding process on the cell image where four-level quantization is performed.

$K = 5$, 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 noted that in both examples biases and classification errors occur because of possible heavy overlaps among closer components. This means that a shift of mean values and 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 they are used for image quantification and segmentation. In our recent work,[4] we developed a framework by combining the 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, so that the convergence rate is increased and the likelihood of being trapped in local optima is reduced.

Table 1 Summary of histogram quantization results (cell image).

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	

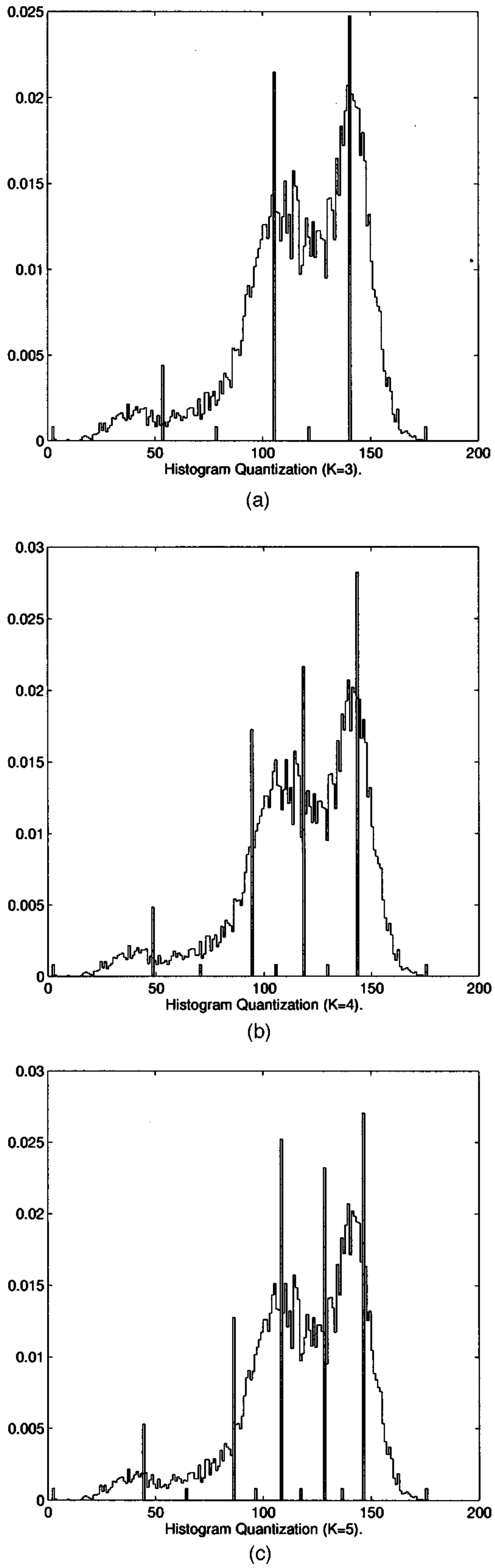


Fig. 7 Results of the optimal quantization of a given histogram (brain), where (a), (b), and (c) correspond to the three-, four-, and five-level quantization, respectively. Similarly, the higher bins in the figures represent the centroids and the lower bins represent thresholds.

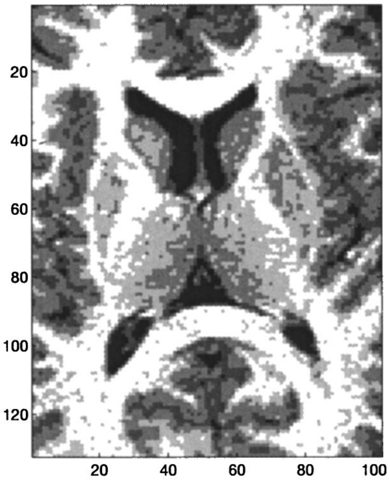


Fig. 8 Result of a simple thresholding process on the MR brain image where five-level (determined by the information theoretic criteria) quantization is performed.

Table 2 summarizes the comparative effects of the initializations by random selection and by SLMHQ scheme on the final quantification and segmentation of an MR brain image. In the quantification experiment, we used a standard finite normal mixture (SFNM) to model the true pixel density distribution and apply the EM algorithm to obtain the maximum likelihood estimate.[3,12] 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 for the GRE value, we could run the EM algorithm with different random initializations. The mean value of the iterations required by EM to reach the specified GRE in 20 independent runs was 67, while when using SLMHQ initialization, only 35 iterations were needed to achieve the same accuracy. Furthermore, based on the initial thresholding result, we used the CBRL algorithm to obtain final contextual segmentation.[4]

Our test shows that, at the stationary point (no pixel relabeling 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.

Table 2 Comparison of random and SLMHQ initializations (MR image).

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 3 Comparison of SLMHQ/SOM/CM (cell image).

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

Second, 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 with evidence that SLMHQ initialization can reduce the likelihood of the solution being trapped in local minima.

We also conducted a comparison study between the SLMHQ selection and Kohonen's self-organizing map (SOM)[9] and the classification-maximization (CM)[3] algorithm since these two methods have also been used frequently to initialize image analysis algorithms in many applications.[9] The evaluation criterion is a critical issue in our comparison since there is no gold standard. In this work, we used both the quantization error (MSE in the gray-level domain) and the quantification error (GRE in the probability domain) as the performance measure. The numerical results are given in Table 3. It can be seen that, in general, the SLMHQ outperforms both SOM and CM algorithms (except for the GRE value of the 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 competitive learning, only the mean difference is taken into account so that the thresholds are the centroids of the means. This may be suitable only when the variances of all components are identical. The CM algorithm uses a modified Mahalanobis distance to achieve a maximum likelihood classification, which clearly improves the final results. However, since the prior probability of each component (e.g., prior in Bayesian classifier) is missing in the formulation, the method cannot deal with the unbalanced mixture cases. In contrast, the results of SLMHQ selection are closest to the Bayesian classification when the image histogram can be modeled by an SFNM distribution.[4,7] In addition, note that neither SOM nor CM is an automatic method since each one still needs an initialization step, which is eliminated in our SLMHQ approach.

4 Conclusion and Extended Work

In this paper, we present an automatic threshold selection method for image analysis, and demonstrate the efficient and reliable application of the algorithm. The technique is unique in that it poses the problem as an optimal scalar quantization problem 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 needed to properly describe the dominant structures of the image at a given number of levels. The method has great promise for 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 and 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 expert neural networks in computer-aided diagnosis[10] where the feature space is two-dimensional. The preliminary results are very satisfactory.[4,10]

Acknowledgment

The authors would like to thank Professor Tianhu Lei for his constructive comments, which led us 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 Vis. Graphics Image Proc.* **41**, 233–260 (1988).
2. M. I. Sezan, ''A peak detection algorithm and its application to histogram-based image data reduction,'' *Computer Vis. Graphics Image Proc.* **49**, 36–51 (1990).
3. Y. Wang and T. Lei, ''A new stochastic model-based image segmentation technique for MR images,'' *Proc. 1st IEEE Intl. Conf. Image Proc.*, pp. 182–185 (1994).
4. Y. Wang, T. Adali, M. T. Freedman, and S. K. Mun, ''MR brain image analysis by distribution learning and relaxation labeling,'' in *Proc. 15th Southern Biomed. Eng. Conf.*, pp. 133–136 (1996).
5. J. C. Olivo, ''Automatic threshold selection using the wavelet transform,'' *Graphical Models Image Proc.* **56**(3), 205–218 (1994).
6. K. V. Mardia and T. J. Hainsworth, ''A spatial thresholding method for image segmentation,'' *IEEE Trans. Pattern Anal. Machine Intell.* **10**(6), 919–927 (1988).
7. T. Kurita, N. Otsu, and N. Abdelmalek, ''Maximum likelihood thresholding based on population mixture models,'' *Pattern Recog.* **25**(10), 1231–1240 (1992).
8. 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).
9. 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, Cambridge, MA (1993).
10. H. Li, S. C. Lo, Y. Wang, W. Hayes, M. T. Freedman, and S. K. Mun, ''Detection of masses on mammograms using advanced segmentation techniques and an HMOE classifier,'' in *Proc. 3rd Int. Workshop on Digital Mammography*, Chicago, pp. 397–400 (1996).
11. J. Max, ''Quantizing for minimum distortion,'' *IRE Trans. Inform. Theory* **6**, 7–12 (1960).
12. Y. Wang, T. Adali, and C. Lau, ''Quantification of MR brain images by probabilistic self-organizing map,'' *Radiology* **197**(P), 252–253 (1995).