Correspondence

A New Criterion for Automatic Multilevel Thresholding

Jui-Cheng Yen, Fu-Juay Chang, and Shyang Chang

Abstract—**In this correspondence, a new criterion for multilevel thresholding is proposed. The criterion is based on the consideration of two factors. The first one is the discrepancy between the thresholded and original images and the second one is the number of bits required to represent the thresholded image. Based on a new maximum correlation criterion for bilevel thresholding, the discrepancy is defined and then a cost function that takes both factors into account is proposed for multilevel thresholding. By minimizing the cost function, the classification number that the gray-levels should be classified and the threshold values can be determined automatically. In addition, the cost function is proven to possess a unique minimum under very mild conditions. Computational analyses indicate that the number of required mathematical operations in the implementation of our algorithm is much less than that of maximum entropy criterion. Finally, simulation results are included to demonstrate their effectiveness.**

I. Introduction

In image processing, it is often necessary to extract objects from an image and represent them efficiently. The most commonly used method is to identify the different homogeneous regions of an image by gray-level thresholding [1]–[3]. The results can then be applied to automatic target recognition [4], text enhancement [5], [6], industrial application of computer vision [7], biomedical image analysis [6], [8], [9], etc. In order to evaluate the threshold values, many methods have been proposed. There are essentially two main approaches: parametric [10], [11] and nonparametric [12], [13]. In the parametric approach, the computational complexity is high. Moreover, the deviation between the histogram of the acquired image and the assumed model usually results in poor performance. On the other hand, the nonparametric approach determines the threshold values in an optimal fashion based on a given criterion [12]–[15]. It can be shown that it is robust and more accurate than the parametric one. However, this approach has two main drawbacks. 1) The classification number that the gray-levels should be classified is difficult to decide and usually is given by supervision. 2) Large computational time is required for determining the threshold values in multilevel thresholding.

In order to overcome these problems, a new criterion for **multilevel** thresholding will be proposed in this correspondence. Two important factors [16] are considered in our approach. The first one is the discrepancy between the thresholded and original images, and the second one is the number of bits required to represent the thresholded image. Based on a new maximum correlation criterion [17] for **bilevel** thresholding, the discrepancy is defined and then a cost function that takes both factors into account is proposed for multilevel thresholding. By minimizing the cost function, the classification number that the gray-levels should be classified and the threshold values can be determined automatically. In addition, the cost function is proved to possess a unique minimum under very mild conditions. The novelty of this criterion is twofold. 1) The classification number and threshold values are automatically determined. 2) The computational complexity is low. Computational analyses and simulation results will demonstrate its effectiveness.

This correspondence is organized as follows. In Section II, the conventional maximum entropy criterion is reviewed and then the maximum correlation criterion is proposed for bilevel thresholding. In Section III, based on the maximum correlation criterion, a new criterion for multilevel thresholding is proposed. Moreover, detailed analysis will also be given. In Section IV, computational analyses and simulation results are given to demonstrate the effectiveness of the new criterion. Section V concludes the correspondence.

Manuscript received April 22, 1992; revised September 4, 1993. The associate editor coordinating the review of this paper and approving it for publication was Prof. Roland T. Chin.

The authors are with the Institute of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, Republic of China.

IEEE Log Number 9408207.

II. The Maximum Correlation Criterion for Bilevel Thresholding

Consider an image $f(x, y)$ of size $N \times N$ pixels that are represented by m gray-levels. Let $\mathbf{G_m} \equiv \{0, 1, \ldots, (m-1)\}$ denote the set of gray-levels and f_i, $i \in \mathbf{G_m}$ be the observed gray-level frequencies of the image f. The probability of the gray-level i in the image f can be calculated as

$$p_i = \frac{f_i}{N \times N}, \quad i \in \mathbf{G_m}.$$

Hence, a distribution $\{p_i \mid i \in \mathbf{G_m}\}$ can be obtained. For a given gray-level s, if $\sum_{i=0}^{s-1} p_i$ is larger than zero and smaller than one, then the following two distributions can be derived from this distribution after normalization:

$$A \equiv \left\{ \frac{p_0}{P(s)}, \frac{p_1}{P(s)}, \ldots, \frac{p_{s-1}}{P(s)} \right\}$$

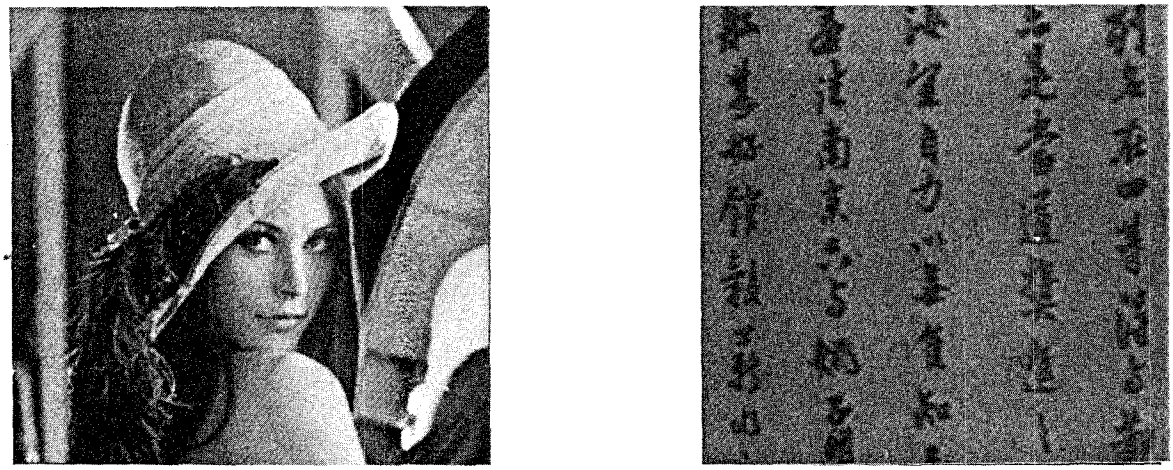
$$B \equiv \left\{ \frac{p_s}{1-P(s)}, \frac{p_{s+1}}{1-P(s)}, \ldots, \frac{p_{m-1}}{1-P(s)} \right\}$$

where $P(s) = \sum_{i=0}^{s-1} p_i$ is the total probability up to the $(s-1)$-th gray-level. In the maximum entropy criterion, the basic idea is to choose the threshold such that the total amount of information provided by the object and background is maximized. Since the information is measured by entropy [18], [19], the total amount of information provided by A and B is

$$\begin{aligned} \mathrm{TE}(s) &= E_A(s) + E_B(s) \\ &= -\sum_{i=0}^{s-1} \left(\frac{p_i}{P(s)} \right) \ln \left(\frac{p_i}{P(s)} \right) \\ &\quad - \sum_{i=s}^{m-1} \left(\frac{p_i}{1-P(s)} \right) \ln \left(\frac{p_i}{1-P(s)} \right) \qquad (1) \\ &= \ln[P(s)(1-P(s))] - H(s)/P(s) \\ &\quad - H'(s)/(1-P(s)) \qquad (2) \end{aligned}$$

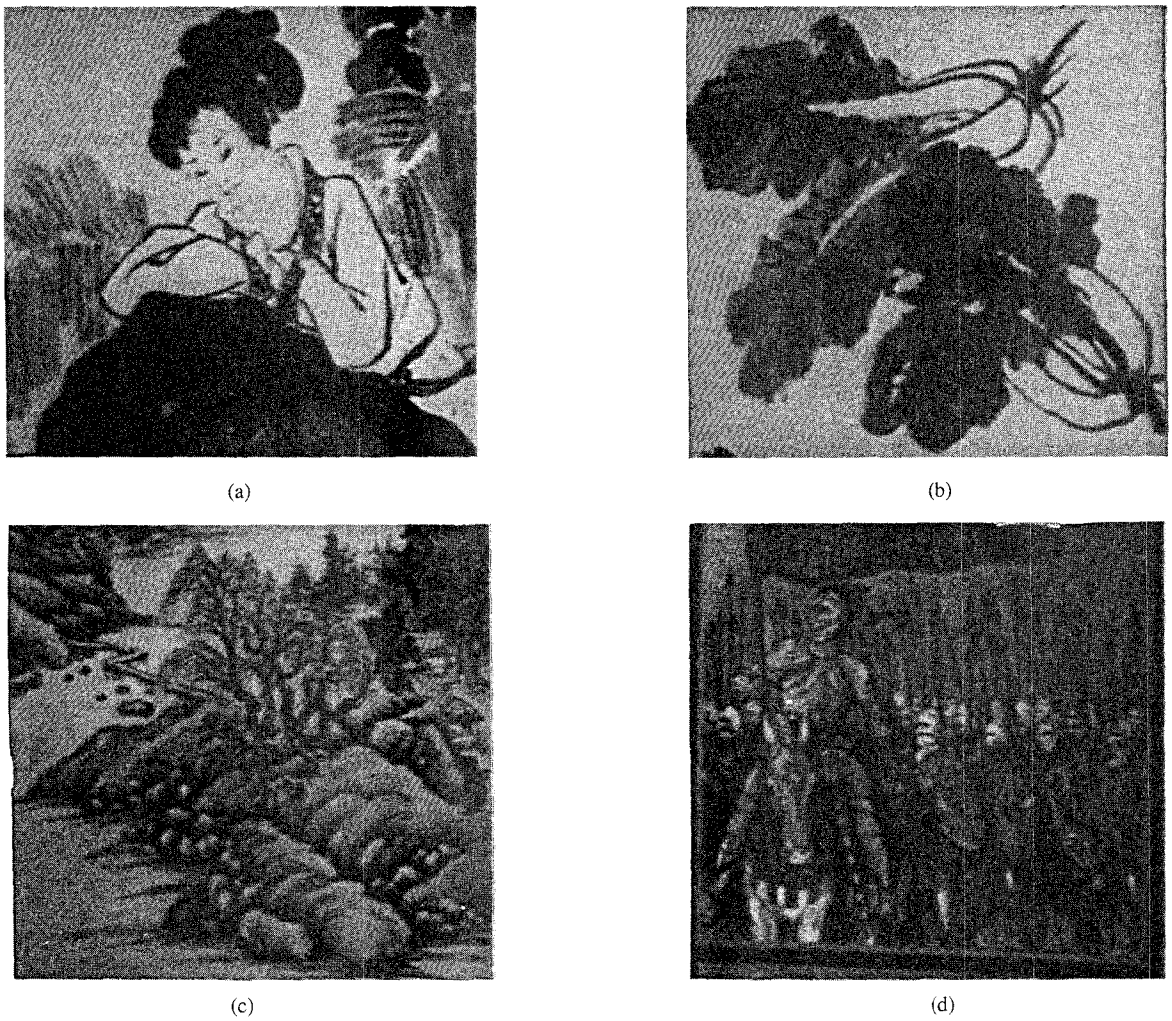
where $P(s) = \sum_{i=0}^{s-1} p_i$, $H(s) = -\sum_{i=0}^{s-1} p_i \times \ln(p_i)$, $H'(s) = -\sum_{i=s}^{m-1} p_i \times \ln(p_i)$, and ln is the natural logarithm. The **maximum entropy criterion (MEC)** [12] is to determine the threshold s^* such that

$$\mathrm{TE}(s^*) = \max_{s \in \mathbf{G_m}} \mathrm{TE}(s). \qquad (3)$$



(a) (b)

Fig. 1. Original images of (a) "LENA" and (b) "Calligraphy" for bilevel thresholding.



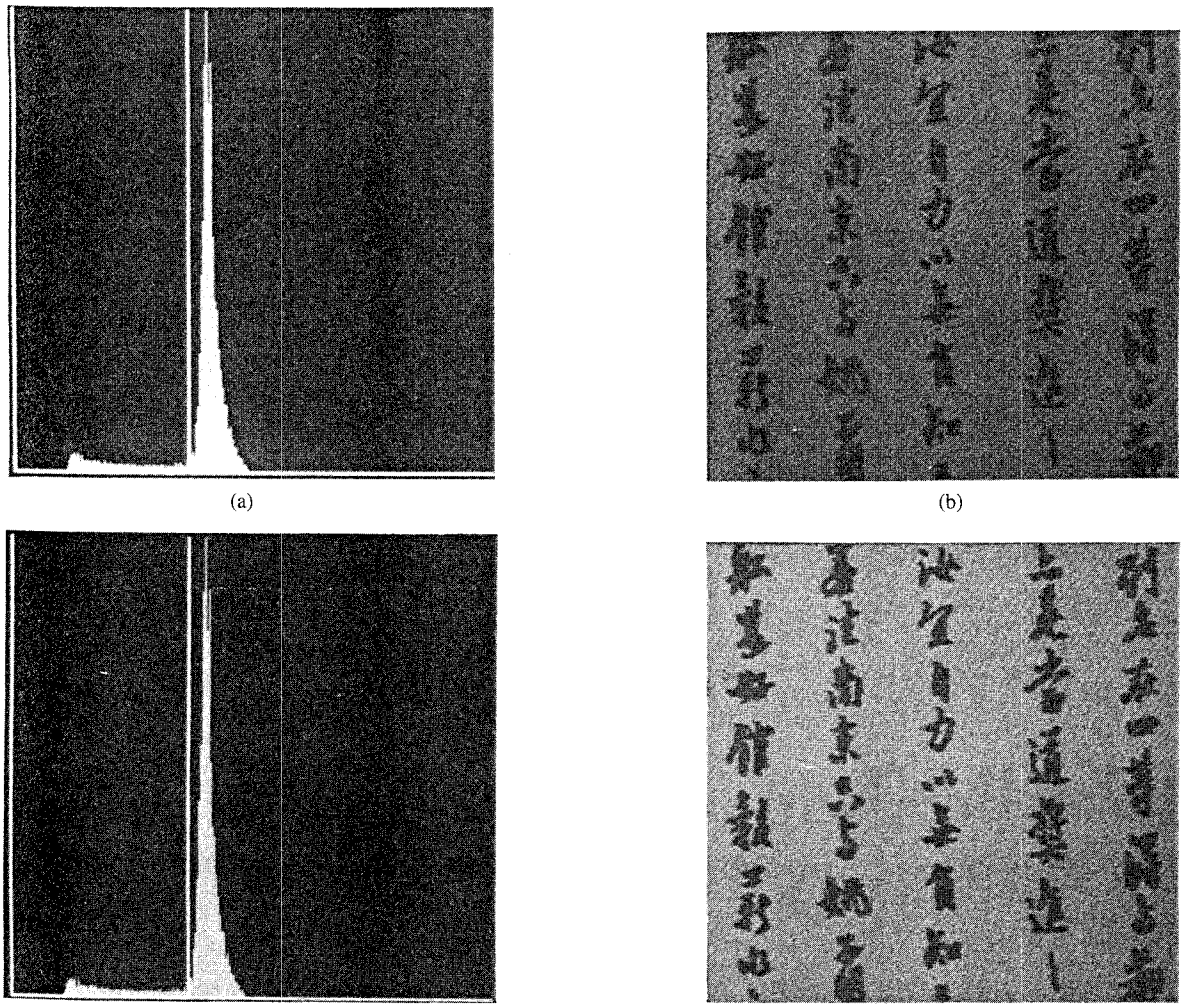
(a) (b)

(c) (d)

Fig. 2. Original images of (a) "Lady," (b) "Vegetable," (c) "Chinese landscape," and (d) "C-Bible" for multilevel thresholding.

The total information defined in (1) instead of the conventional average information $P(s)E_A(s)+(1-P(s))E_B(s)$ has been adopted by [12]. The reason is that, if the conventional one is used, the object(s) with low priors $P(s)$ or $(1-P(s))$ cannot be segmented by the MEC and eventually disappear in the thresholded image.

Notice that in (1) and (3), many natural logarithmic operations are required for bilevel thresholding via MEC. In order to reduce the computational complexity, the maximum correlation criterion will be proposed. Before proposing it, the correlation [17], [20] will be defined as follows.



(a) (b) (c) (d)

Fig. 3. Gray-level histograms and the optimal thresholded positions of "Calligraphy" via (a) MCC and (c) MCC. The thresholded images of "Calligraphy" via (b) MCC and (d) MEC.

TABLE I
NUMBERS OF DIFFERENT KINDS OF OPERATIONS REQUIRED BY MCC AND MEC

equation	operations				
	ln	÷	×	+ or -	<>
equation (4)	1	1	(m+5)	<(3m/2)	0
equation (5)	2	0	(m+3)	<(3m/2+1)	0
MCC eq. (6)	m	m	m(m+5)	<m(3m/2)	m-1
equation (1)	m	m	m	<(3m/2)	0
equation (2)	m+1	2	m+1	<(3m/2+4)	0
MEC eq. (3)	m(m+1)	2m	m(m+1)	<m(3m/2+4)	m-1

- "m" denotes the number of gray-levels used to represent an image.
- "<(3m/2)" denotes the required operations are less than 3m/2.
- "<>" denotes the comparison operation.

TABLE II
OPTIMAL THRESHOLD VALUES LOCATED BY MCC AND MEC FOR "LENA" AND "CALLIGRAPHY"

criterion \ image	LENA	Calligraphy
MCC	120	94
MEC	122	95

Definition: Let $\mathbf{X}$ be a discrete random variable with finite or countably infinite range $R = \{x_0, x_1, x_2, \ldots\}$ and p_i denote Prob$\{\mathbf{X} = x_i\}$. The correlation of $\mathbf{X}$ is defined as

$$C(\mathbf{X}) = -\ln \sum_{i \geq 0} p_i^2.$$

The idea of using correlation $C(s)$ instead of entropy $E(s)$ comes from the theory of chaos and fractal. In fractal theory, both entropy dimension and correlation dimension are used for image representation and real objects modeling. However, the computational complexity of computing correlation dimension is much lower than that of the entropy dimension. Hence, the idea of correlation is adopted here.

Based on the definition, the total amount of correlation provided by distributions A and B is

$$\begin{aligned} \mathrm{TC}(s) &= C_A(s) + C_B(s) \\ &= -\ln \sum_{i=0}^{s-1} \left(\frac{p_i}{P(s)}\right)^2 - \ln \sum_{i=s}^{m-1} \left(\frac{p_i}{1-P(s)}\right)^2 \\ &= -\ln\{[G(s) \times G'(s)]/[P^2(s) \times (1-P(s))^2]\} \end{aligned} \quad (4)$$

TABLE III
CLASSIFICATION NUMBER, THE THRESHOLD VALUES, AND THE REQUIRED CPU TIMES FOR "LADY," "VEGETABLE," "CHINESE LANDSCAPE," AND "C-BIBLE" VIA ATC AND MEC

image	criterion	class. no.	threshold values	CPU time
Lady	ATC	4	(1)95 (2)55 (3)121	0.11"
	MEC	4*	55, 85, 116	224.25"
Vegetable	ATC	3	(1)86 (2)58	0.12"
	MEC	3*	64, 102	4.07"
Chinese landscape	ATC	5	(1)85 (2)115 (3)61 (4)100	0.10"
	MEC	5*	56, 75, 95, 115	3346.45"
C-Bible	ATC	3	(1)83 (2)56	0.09"
	MEC	3*	62, 102	4.29"

• "*" denotes the classification number is given by supervision.

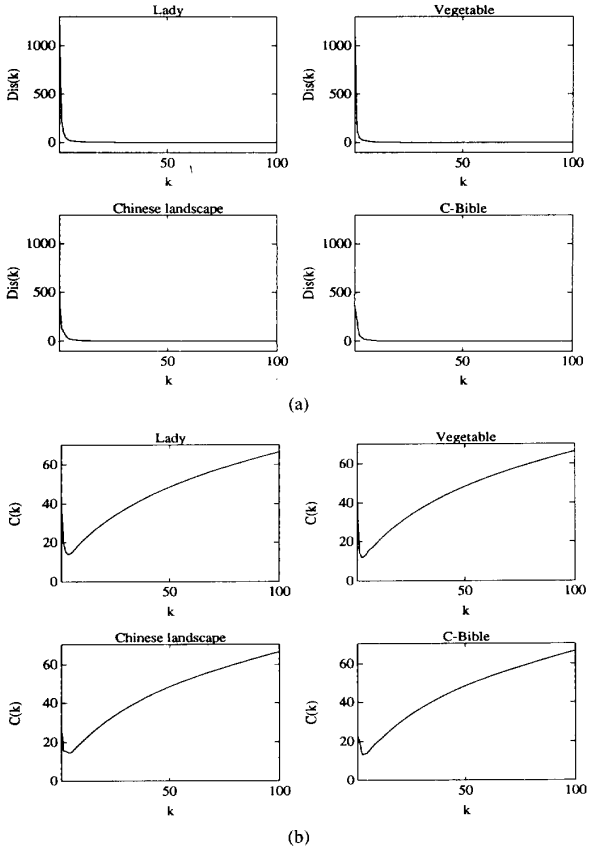


Fig. 4. Functions (a) Dis(k) and (b) $\mathbf{C}(k)$ of four representative images by ATC.

$$= -\ln[G(s) \times G'(s)] + 2\ln[P(s) \times (1 - P(s))] \quad (5)$$

where $G(s) = \sum_{i=0}^{s-1} p_i^2$ and $G'(s) = \sum_{i=s}^{m-1} p_i^2$. In order to obtain the maximal correlation contributed by the object and background in the image f, TC(s) must be maximized. The **maximum correlation criterion (MCC)** is to determine the threshold s^* such that

$$\mathrm{TC}(s^*) = \max_{s \in \mathbf{G_m}} \mathrm{TC}(s). \quad (6)$$

III. AUTOMATIC MULTILEVEL THRESHOLDING

It is well-known that the thresholded image becomes more similar to the original one as the classification number increases. Hence, the discrepancy between the original and thresholded images decreases as the classification number increases. However, the total number of bits required to represent the thresholded image increases as the number of classes increases. Hence, there must exist a compromise between these two factors.

Let k denote the classification number and Dis(k) the discrepancy between the thresholded and original images. The cost function $\mathbf{C}(\bullet)$ that takes into account both factors is proposed as

$$\mathbf{C}(k) = \rho(\mathrm{Dis}(k))^{1/2} + (\log_2(k))^2 \quad (7)$$

where ρ is a positive weighting constant. The first term of $\mathbf{C}(k)$ measures the cost incurred by the discrepancy between the thresholded and original images, and the second measures the cost resulted from the number of bits used to represent the thresholded image. The square root of discrepancy Dis(k) has a similar meaning to the "standard deviation" of a random variable. There are two reasons why we adopt the square bit-usage penalty. The first is to use the "intermediate value theorem" in the proof that $\mathbf{C}(k)$ can possess a unique minimum. The second is to avoid the domination of this term by $(\mathrm{Dis}(k))^{1/2}$. To achieve the best compromise between $\mathbf{C}_d(\mathrm{Dis}(k))$ and $\mathbf{C}_b(\log_2(k))$, $\mathbf{C}(k)$ must be minimized. Based on the cost function $\mathbf{C}(k)$, the **automatic thresholding criterion (ATC)** is then proposed to determine the optimal classification number k^* such that

$$\mathbf{C}(k^*) = \min_{k \in \mathbf{G}_\mathbf{m}^+} \mathbf{C}(k) \quad (8)$$

where $\mathbf{G}_\mathbf{m}^+ \equiv \{1, 2, \ldots, m\}$.

In order to quantify Dis(k), the following notations will be adopted:

- k the classification number that the gray-levels are classified,
- $s_{k,i}$ the i-th nonzero threshold when the gray-levels are classified into k classes,
- $C_{k,i}$ the i-th class among these k classes with gray-levels from $s_{k,i-1}$ to $(s_{k,i} - 1)$,
- $\omega_{k,i}$ the probability of the class $C_{k,i}$,
- $P_{k,i}$ the distribution derived from $C_{k,i}$ after normalized by $\omega_{k,i}$,
- $\mu_{k,i}$ the mean of $P_{k,i}$,
- $\sigma_{k,i}^2$ the variance of $P_{k,i}$.

Hence, for a given distribution $P \equiv \{p_i \mid i \in \mathbf{G_m}\}$, the following relations can be obtained:

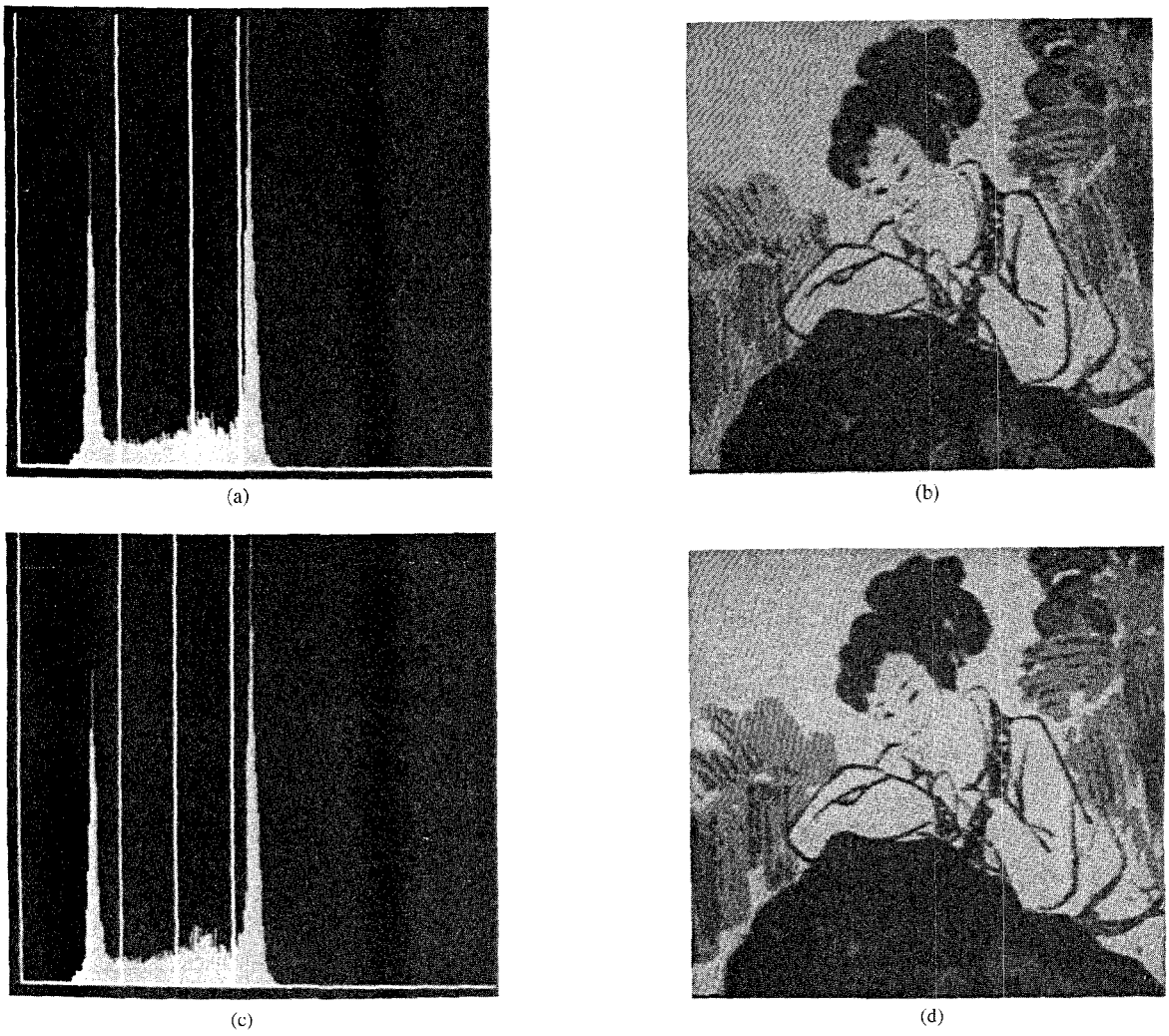
$$\omega_{k,i} = \mathrm{Prob}(C_{k,i}) = \sum_{j=s_{k,i-1}}^{s_{k,i}-1} p_j \quad (9)$$

$$P_{k,i} = \{p_j/\omega_{k,i} \mid j \in \mathbf{G}_{\mathbf{s}_{k,i}}\} \quad (10)$$

$$\mu_{k,i} = \sum_{j=s_{k,i-1}}^{s_{k,i}-1} j \times \mathrm{Prob}(j \mid C_{k,i}) \quad (11)$$

$$\sigma_{k,i}^2 = \sum_{j=s_{k,i-1}}^{s_{k,i}-1} (j - \mu_{k,i})^2 \times \mathrm{Prob}(j \mid C_{k,i}) = \sum_{j=s_{k,i-1}}^{s_{k,i}-1} (j - \mu_{k,i})^2 \times p_j/\omega_{k,i} \quad (12)$$

where $\mathbf{G}_{s_{k,i}} \equiv \{s_{k,i-1}, s_{k,i-1} + 1, \ldots, (s_{k,i} - 1)\}$. In each class, all the gray-levels are designated to the mean value of the gray-levels



(a) (b) (c) (d)

Fig. 5. Gray-level histograms and the optimal thresholded positionsof "Lady" via (a) ATC and (c) MEC. The thresholded images of "Lady" via (b) ATC and (d) MEC.

in that class. The discrepancy Dis(k) is then defined as

$$\begin{aligned}\mathrm{Dis}(k) &= \sum_{j=1}^{k} \mathrm{Prob}(C_{k,j}) \times \sigma_{k,j}^2 \\ &= \sum_{i=0}^{s_{k,1}-1} (i-\mu_{k,1})^2 \times p_i \\ &\quad + \sum_{i=s_{k,1}}^{s_{k,2}-1} (i-\mu_{k,2})^2 \times p_i \\ &\quad + \cdots + \sum_{i=s_{k,k-1}}^{m-1} (i-\mu_{k,k})^2 \times p_i. \end{aligned} \tag{13}$$

Choosing the distribution with the largest variance from $P_{k,i}$'s and applying MCC to it, this selected distribution can be further dichotomized into two more distributions. Hence, the original distribution P can be partitioned into $(k+1)$ distributions after normalization and the gray-levels of the original image are also divided into $(k+1)$ classes. Relabel all the thresholds and the $(k+1)$-class version of (9)–(12) can be obtained. Hence, the discrepancy Dis$(k+1)$ can be defined accordingly.

Intuitively, the thresholded image becomes more similar to the original one as the classification number increases. Hence, the discrepancy should decrease when the classification number increases. Now, the defined Dis(k) will be proved to possess this property in the following.

Proposition 1: Dis(k) is strictly decreasing to zero for $k \in \mathbf{G}_{\mathbf{m}}^{\div}$.

Proof: Straightforward. □

Furthermore, if we assume Dis(k) is of the form $\alpha \times k^{-\lambda}$, then the following proposition concerning $\mathbf{C}(k)$ can be obtained.

Proposition 2: If $\mathrm{Dis}(k) = \alpha \times k^{-\lambda}$, $\alpha > 0$, $\lambda > 0$, and $\rho < 4 \times \ln 256 \times 256^{\lambda/2}/(\lambda \alpha^{1/2} (\ln 2)^2)$, then $\mathbf{C}(k)$ has a unique minimum for $k \in [1, 256]$.

Proof: It is straightforward by using the intermediate value theorem. □

From Proposition 2, instead of searching in $\mathbf{G}_{\mathbf{m}}^{+}$ globally, the minimum cost can be determined when $\mathbf{C}(k)$ starts to increase. This implies that the searching procedure in (8) can be terminated and the computational complexity of ATC can be further reduced.

IV. Computational Analyses and Simulation Results

To demonstrate the effectiveness of the proposed MCC and ATC for bilevel and multilevel thresholdings, computational analyses and

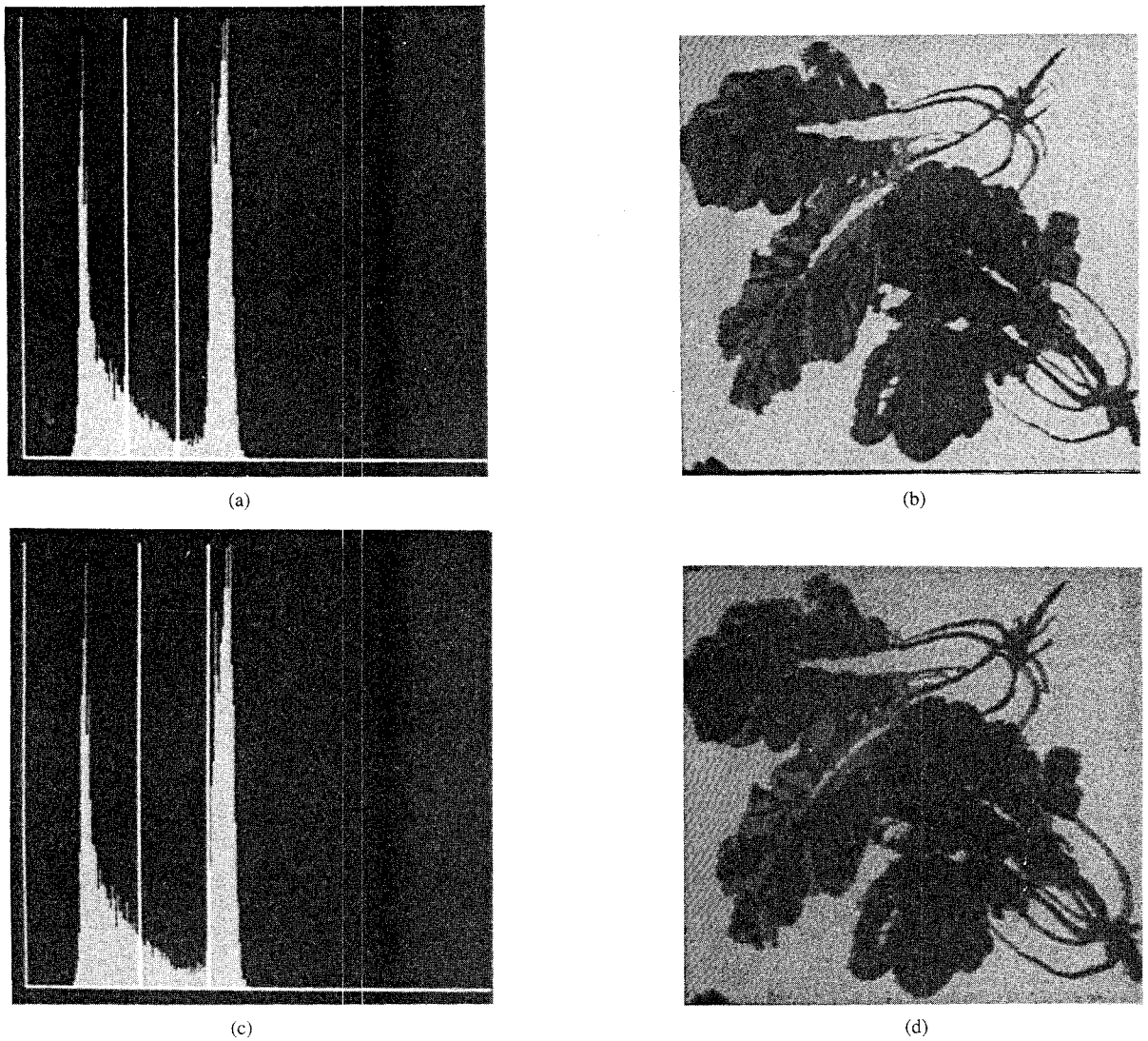


Fig. 6. Gray-level histograms and the optimal thresholded positions of "Vegetable" via (a) ATC and (c) MEC. The thresholded images of "Vegetable" via (b) ATC and (d) MEC.

simulation results on the HP 9000/720 workstation with 17 MFLOPS will be given. For bilevel thresholding, there are two images of size 256 × 256, called "LENA" and "Calligraphy." To compare the multilevel thresholdings via ATC and MEC, four representative images of size 256 × 256, called "Lady," "Vegetable," "Chinese landscape," and "C-Bible," are used. They are shown in Figs. 1 and 2.

A. Comparisons of Bilevel Thresholdings between MEC and MCC

The numbers of the required operations for MCC and MEC are listed in Table I. The optimal thresholds located by both criteria for these two images are listed in Table II. As a representative, Fig. 3(a) and (c) illustrates the gray-level histograms combined with the optimal threshold positions for "Calligraphy" via MCC and MEC, respectively. All the gray-levels in each class are designated to the mean value of the gray-levels in that class. Figure 3(b) and (d) shows the thresholded images of "Calligraphy" via MCC and MEC, respectively.

From Table I, the number of the natural logarithmic operations is drastically reduced from $(m^2 + m)$ for MEC to m for MCC. Moreover, the number of divisions is reduced from $2m$ for MEC to m for MCC. The difference in the number of the other operations is very small. From Table II and Fig. 3(a) and (c), the optimal thresholds for "Calligraphy" via MCC and MEC are almost the same. Hence, the quality of the thresholded images cannot be distinguished by human eyes from Fig. 3(b) and (d).

These results indicate that the computational complexity of MCC is drastically reduced as compared to MEC, while the thresholded images are almost the same.

B. Comparisons of Multilevel Thresholdings between ATC and MEC

In order to apply ATC to these images, ρ is determined by a set of test images. In this section, ρ is 0.8. The classification number that the gray-levels are classified, the threshold values, and the required CPU times are all listed in Table III. The functions $\mathrm{Dis}(k)$ and $\mathbf{C}(k)$ of "Lady," "Vegetable," "Chinese landscape," and "C-Bible" are plotted in Fig. 4(a) and (b), respectively. The gray-level histograms combined with its optimal thresholded positions are shown in Figs. 5(a), 6(a), 7(a), and 8(a). After the optimal thresholds have been located for each image, we designate all the gray-levels in each class to the mean value of the gray-levels in that class. Then, the thresholded images are shown in Figs. 5(b), 6(b), 7(b), and 8(b).

In using the MEC for multilevel thresholding, the classification number k must be given by supervision. That is, the gray-levels of "Lady," "Vegetable," "Chinese landscape," and "C-Bible" must be

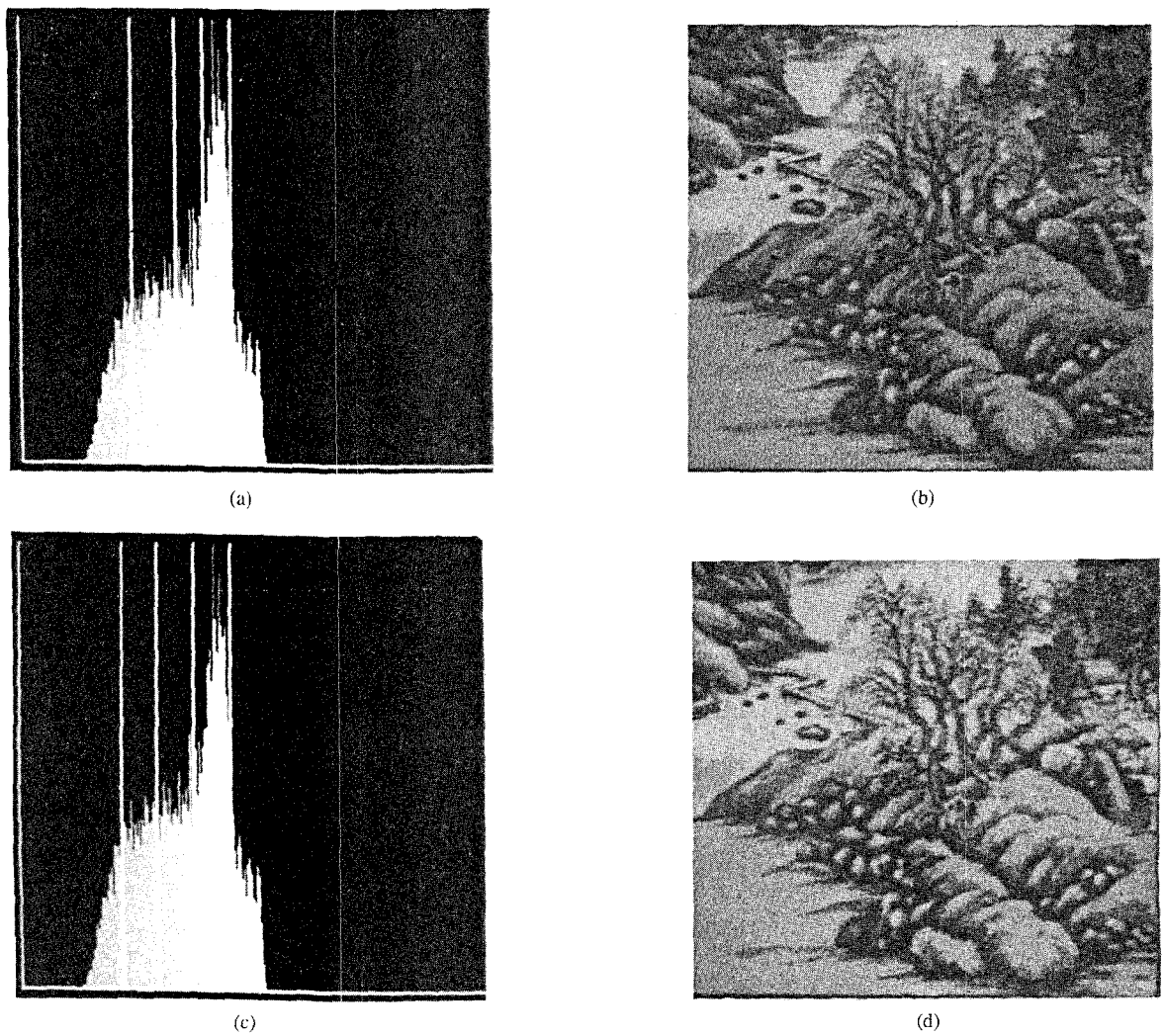


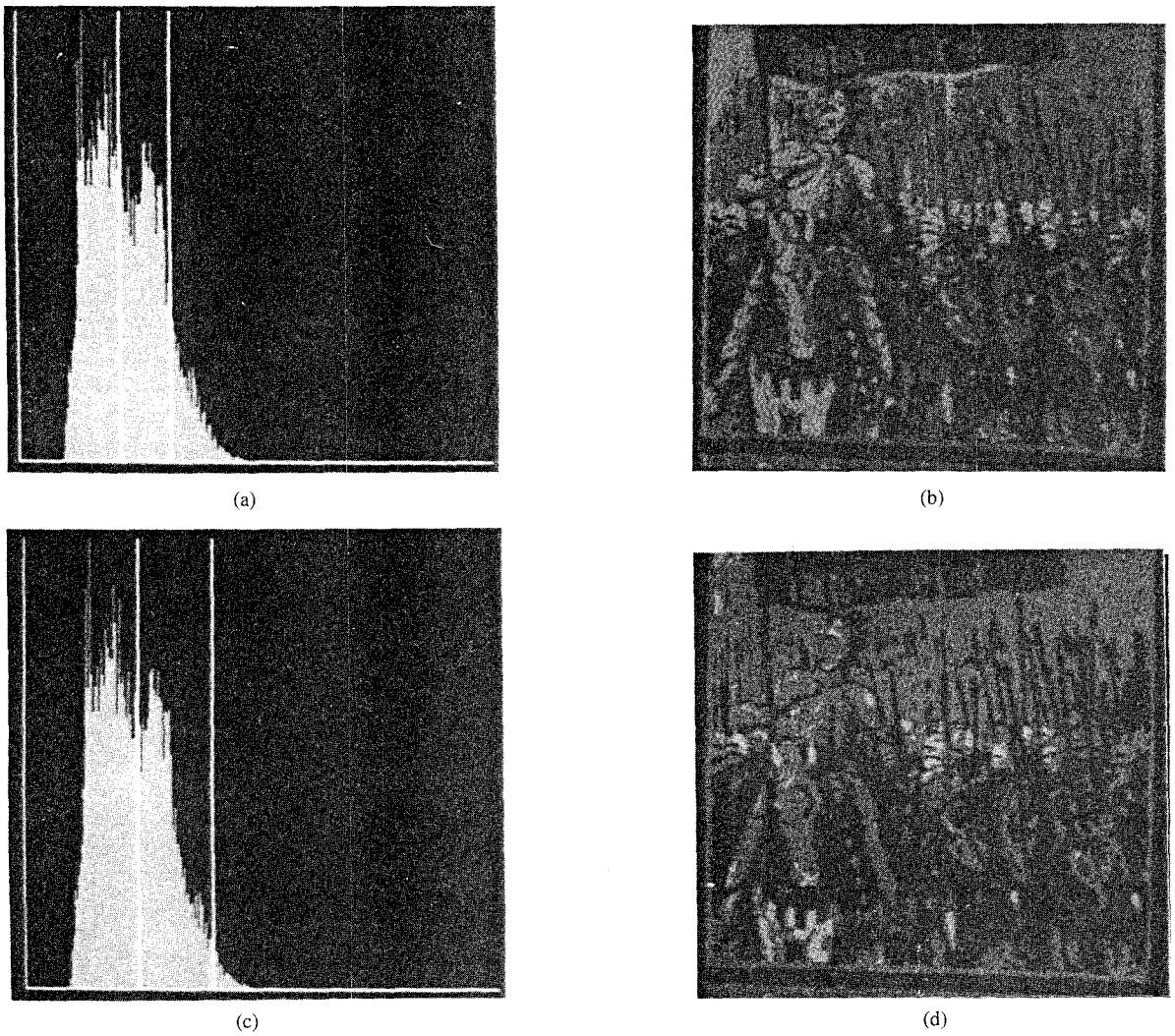
Fig. 7. Gray-level histograms and the optimal thresholds of "Chinese landscape" via (a) ATC and (c) MEC. The thresholded images of "Chinese landscape" via (b) ATC and (d) MEC.

classified into 4, 3, 5, and 3 classes, respectively. After applying MEC to these images, the located optimal thresholds and required CPU times are listed in Table III for comparison. The gray-level histograms combined with their optimal thresholded positions are shown in Figs. 5(c), 6(c), 7(c), and 8(c). After the optimal thresholds are located, all the gray-levels in each class are designated to the mean value of the gray-levels in that class. Then, the thresholded images are shown in Figs. 5(d), 6(d), 7(d), and 8(d).

As can be seen from (14), it is too complex and computationally expensive to minimize the cost function of MEC as a function of thresholds although the classification number k has been given by supervision. On the other hand, our sequential way of finding thresholds via ATC can reduce the complexity drastically. This can be confirmed by the required CPU time in Table III. Moreover, as illustrated in Figs. 5–8, our results are quite good. Hence, compared with MEC, the ATC is a very computationally efficient criterion that has good performance for multilevel thresholding.

In Section III, the discrepancy function $\mathrm{Dis}(k)$ has been proved to be strictly decreasing to zero. This can also be confirmed by the simulation results in this section. After applying ATC to these four images, Fig. 4(a) indicates that $\mathrm{Dis}(k)$ is such a decreasing function indeed. Moreover, it confirms that our assumption, i.e., $\mathrm{Dis}(k)$ is of the form $\alpha \times k^{-\lambda}$, is quite reasonable. In Proposition 2, we have proved that each cost function $\mathbf{C}(k)$ possesses a unique minimum. This fact is also confirmed by the four cases shown in Fig. 4(b). In this figure, $\mathbf{C}(k)$ decreases to the minimal value at some k, and then increases to $\mathbf{C}(256)$. Hence, to find the minimum cost, we just need to calculate $\mathbf{C}(k)$ successively until it starts to increase. If the searching procedure in (8) is performed this way, the required CPU time of ATC listed in Table III can be further reduced.

In the image "Lady" shown in Fig. 2(a), the gray-levels can be coarsely classified into four classes by human eyes. After applying ATC to "Lady,",the image is segmented into four homogeneous regions. As can be seen from Fig. 5(b) and (d), the veins and outline of the face of using ATC are more clear than that of MEC. In the images "Vegetable" and "C-Bible," by using only two thresholds determined by ATC, the thresholded images are very much like the original ones. As can be seen from Fig. 8(b) and (d), the thresholded image of "C-Bible" via ATC has more clear faces and shields than that of MEC. Moreover, in Fig. 6(d), there are many spots in the thresholded image of "Vegetable" via MEC. In the other, more complicated image "Chinese landscape," almost all important components are reserved in the thresholded image. Moreover, the stones' texture in Fig. 7(b) is more clear than that in Fig. 7(d). The results above indicate that the ATC can automatically segment these images under the compromise between the discrepancy and the



(a) (b) (c) (d)

Fig. 8. Gray-level histograms and the optimal thresholded positions of "C-Bible" via (a) ATC and (c) MEC. The thresholded images of "C-Bible" via (b) ATC and (d) MEC.

classification number. Moreover, the thresholded results of ATC are better than those of MEC. These are confirmed by the simulation results shown in Table III and Figs. 5(b), 5(d), 6(b), 6(d), 7(b), 7(d), 8(b), and 8(d).

Notice that if the gray-level histogram and the spatial correlation between the pixels are both considered in an image, then the 1-D histogram can be extended to a 2-D one. In this case, the MCC and ATC can also be extended to deal with a 2-D histogram for bilevel and multilevel thresholdings and preserve the desirable properties indicated above.

V. CONCLUSION

In this correspondence, the MCC and ATC have been proposed for multilevel thresholding to overcome the drawbacks of conventional approaches. The cost function has been proven to possess a unique minimum under very mild conditions. This proposition has reduced the searching procedure of ATC and its computational complexity. Moreover, the following desirable properties have been obtained. 1) The number of required mathematical operations in the implementation of MCC is much less than that of MEC. 2) The classification number that the gray-levels should be classified and those threshold values via ATC are automatically determined. 3) The computational complexity of ATC is much less than MEC. Simulation results have also confirmed the aforementioned properties. We believe that many applications in image processing can benefit from using the ATC.

REFERENCES

[1] J. S. Weszka, "A survey of threshold selection techniques," *Comput. Graph. Image Processing*, vol. 7, no. 2, pp. 259–265, Apr. 1978.

[2] K. S. Fu and J. K. Mui, "A survey on image segmentation," *Pattern Recogn.*, vol. 13, no. 1, pp. 3–16, 1981.

[3] P. K. Sahoo, S. Soltani, A. K. C. Wong, and Y. C. Chen, "A survey of thresholding techniques," *Comput. Vision, Graph., Image Processing*, vol. 41, no. 2, pp. 233–260, Feb. 1988.

[4] B. Bhanu, "Automatic target recognition: State of the art survey," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 4, pp. 364–379, July 1986.

[5] T. W. Ridler and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Trans. Syst. Man Cyber.*, vol. SMC-8, no. 8, pp. 630–632, Aug. 1978.

[6] J. R. Parker, "Gray-level thresholding in badly illuminated images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 8, pp. 813–819, Aug. 1991.

[7] R. C. Gonzalez and R. Safabakhsh, "Computer vision techniques for industrial applications and robot control," *Comput.*, vol. 15, pp. 17–32, 1982.

[8] C. K. Chow and T. Kaneko, "Automatic boundary detection of left ventricle from cineangiograms," *Comput. Biomed. Res.*, vol. 5, pp. 338–410, 1972.

[9] E. S. Gelsema, H. F. Bao, A. W. M. Smeulders, and H. C. D. Harink, "Application of the method of multiple thresholding to white blood cell classification," *Comput. Biol. Med.*, vol. 18, no. 2, pp. 65–74, 1988.
[10] J. S. Weszka and A. Rosenfeld, "Histogram modification for threshold selection," *IEEE Trans. Syst. Man Cyber.*, vol. SMC-9, no. 1, pp. 38–52, Jan. 1979.
[11] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. New York: Academic, 1982, vol. 2.
[12] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Comput. Vision, Graph., Image Processing*, vol. 29, no. 3, pp. 273–285, Mar. 1985.
[13] A. S. Abutaleb, "Automatic thresholding of gray-level pictures using two-dimensional entropy," *Comput. Vision, Graph., Image Processing*, vol. 47, no. 1, pp. 22–32, July 1989.
[14] S. S. Reddi, S. F. Rudin, and H. R. Keshavan, "An optimal multiple threshold scheme for image segmentation," *IEEE Trans. Syst. Man Cyber.*, vol. SMC-14, no. 4, pp. 661–665, July/Aug. 1984.
[15] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cyber.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
[16] A. M. Bruckstein, "On 'soft' bit allocation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 5, pp. 614–617, May 1987.
[17] F. J. Chang, J. C. Yen, and S. Chang, "Gray-level thresholding via maximum correlation criterion," in *Proc. 3rd Int. Conf. Advances Commun. Contr. Syst.*, Victoria, Canada, 1991.
[18] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: Univ. of Illinois Press,, 1949.
[19] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[20] H. G. Schuster, *Deterministic Chaos: An Introduction*. Physik-Verlag, 1984.

Motion Vector Quantization for Video Coding

Yoon Yung Lee and John W. Woods, *Fellow, IEEE*

***Abstract*—A new algorithm is developed for the vector quantization of motion vectors. This algorithm, which is called motion vector quantization (MVQ), simultaneously estimates and vector quantizes motion vectors by reinterpreting the block matching algorithm as a type of vector quantization. An iterative design algorithm, based on this concept, is then developed. In addition to reducing rate for fixed length encoding, the algorithm also reduces computation considerably. We include promising coding simulation results on the *Flower Garden* sequence.**

I. Introduction

Since forward motion compensation using a block matching algorithm (BMA) requires transmission of motion vectors as side information [6], the block size must be selected as a compromise between the desire to have a small prediction error and the need to control the amount of side information. This correspondence introduces a new combined estimation and vector quantization algorithm for motion vectors, which is called motion vector quantization (MVQ), to reduce the number of motion vectors needed.

Manuscript received July 2, 1993; revised January 21, 1994. This work was supported by NSF grant no. NCR-521105. The associate editor coordinating the review of this paper and approving it for publication was Prof. Nasser M. Nasrabadi.

The authors are with the Center for Image Processing Research, Department of Electrical and Computer Science Engineering, Rensselaer Polytechnic Institute, Troy, NY, 12180–3590 USA.

IEEE Log Number 9407606.

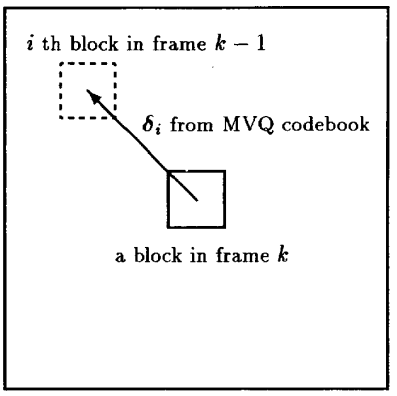


Fig. 1. The relationship between the search region and the codebook.

The most common distortion measure used in waveform coding is the mean squared error (MSE). However, this distortion measure is not appropriate for coding motion vectors because it does not consider the effect of distortion on the displaced frame difference (DFD), defined as the error between the current frame and the motion predicted frame. Our MVQ for motion vectors minimizes the variance of the DFD directly, while iteratively optimizing a codebook of motion vectors.

II. Motion Vector Quantization

In the forward matching, BMA is used for obtaining motion vectors mainly because it requires a small amount of side information. For a given block in the current frame, BMA searches for the best match within a search region in the previous frame. The displacement that represents the best match becomes the motion vector for the block. BMA can be re-interpreted as a type of vector quantization (VQ). The search region can be represented as a codebook, $C = \{\delta_1, \delta_2, \ldots, \delta_N\}$, where the motion vector , $\delta_i = (\delta_{i1}, \delta_{i2})^T$ for $1 \leq i \leq N$, are thought of as code vectors. Fig. 1 shows the relationship between search region and the codebook.

The codebook, or search region, has N motion vectors to represent N possible block locations in the previous frame. The ith code vector δ_i is the motion vector that represents the ith matching block in the search region. The required rate to transmit the motion vectors is then $r = (\log_2 N)/M^2$ for an $M \times M$ block size and fixed length encoding.

Let $s_t(x)$ and $s_{t-1}(x)$ be the current and the previous frames, respectively. We define an input for BMA as a set that consists of a block

$$\bar{s}_t^B = \{s_t(x), x \in \text{ block B}\} \tag{1}$$

within the current frame and a portion

$$\bar{m}_{t-1}^B = \{s_{t-1}(x), x\epsilon \text{ search region for B}\} \tag{2}$$

of the previous frame in the search region. We can define a distortion measure associated with the *input set*, $\{\bar{s}_t^B, \bar{m}_{t-1}^B\}$ and the ith motion