From these two examples, we see that our $P-P$ granule characterization of the transition matrix allows us to make inferences with knowledge structures about V other than simply the pure probability distribution.

One other further benefit of this approach is that it allows us to integrate easily the multiple relations between variables, one of which may be a transition matrix.

Assume V and U are two variables. Assume we have two pieces of information indicating the relationship between these variables. The first is transition matrix T of the type we have just discussed. The second is an implication statement of the form

$$\text{"if } V \text{ is } A \text{ then } U \text{ is } B\text{,"}$$

where A and B are fuzzy subsets of X and Y.

We can represent the first relationship, the transition matrix as a $P-P$ granule

$$(V,U) \text{ is } m_1$$

where m_1 is of the form discussed. The second relationship

$$\text{"if } V \text{ is } A \text{ then } U \text{ is } B\text{"}$$

can also be represented as a P–P granule of the form

$$(V,U) \text{ is } m_2$$

where m_2 has one focal element H where

$$H(x,y) = 1 \wedge (1 - A(x) + B(y))$$

and of course $m_2(H) = 1$.

These two P relations can be conjuncted to give an overall relationship

$$(V,U) \text{ is } m$$

where $m = m_1 \cap m_2$. In this case if $G_k, k = 1, \cdots$ are the focal elements of m_1 then the focal elements of m are $H_k, k = 1 \cdots$, where $H_k = G_k \cap H$ and $m(H_k) = m_1(G_k)$. We could then use this new effective relationship with any information we have about V to make inferences about V.

Conclusion

We have shown how to represent transitional matrices in the framework of P–P granules.

References

[1] R. R. Yager, "Reasoning with uncertainty for expert systems," in *Proc. Ninth IJCAI*, Los Angeles, CA, 1985, pp. 1295–1297.

[2] ____, "Toward a general theory of reasoning with uncertainty—Part I: Nonspecificity and fuzziness," *Int. J. Intelligent Syst.*, vol. 1, pp. 45–67, 1986.

[3] ____, "Toward a general theory of reasoning with uncertainty Part II: Probability," *Int. J. Man–Machine Studies*, (to appear).

[4] L. A. Zadeh, "A theory of approximate reasoning," in *Selected Papers of L. A. Zadeh on Fuzzy Sets and Related Topics*, R. R. Yager, S. Ovchinnikov, R. Tong, and H. Nguyen, Eds. New York: Wiley, 1987.

[5] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.

[6] R. R. Yager, "Arithmetic and other operations on Dempster-Shafer structures," *Int. J. Man–Machine Studies*, vol. 25, pp. 357–366, 1986.

[7] J. Pearl, "Fusion, propagation and structuring in Bayesian networks," Univ. of California, Los Angeles, Tech. Rep. CSD-850022, 1985.

Segmentation Based on Measures of Contrast, Homogeneity, and Region Size

SANKAR K. PAL, SENIOR MEMBER, IEEE, AND NIKHIL R. PAL

***Abstract*—Two algorithms are described for automatic image segmentation using a "homogeneity" measure and "Contrast" measure defined on the cooccurrence matrix of the image. The measure of contrast involves the concept of logarithmic response (adaptibility with background intensity) of the human visual system. Provisions are also kept in two different ways to remove the undesirable thresholds. The effectiveness of the algorithms is demonstrated for a set of images having different types of histograms. The performance of the algorithms is compared to existing ones.**

I. Introduction

One of the key problems in scene analysis is segmentation of a scene into different regions. Segmentation is essentially a pixel classification problem where one tries to classify the pixels into different classes such that each class is homogeneous and at the same time the union of no two adjacent classes is homogeneous. In other words, given a definition of uniformity, segmentation is a partition of the picture into connected subsets, each of which is uniform but such that no union of adjacent subsets is uniform [1].

Several techniques of image segmentation exist, based on global and local information of an image. One of the techniques based on global information is histogram thresholding which selects the valley points as threshold levels. For images where the histogram does not have sharp valleys (i.e., having flat minima or local minima) the histogram is usually sharpened [2]–[6] by a suitable transformation so that the task of selecting valleys becomes easier. These transformations usually require some parameters the choice of which has a significant impact in determining the number of thresholds. The cooccurrence matrix, on the other hand, uses local spatial information of an image and provides information regarding the number of transitions between any two gray levels in the image. This information has been used by various authors, namely, Weszka and Rosenfeld [7], Deravi and Pal [8], and Chanda *et al.* [9], for segmentation.

The measures on the cooccurrence matrix reported by these authors did not consider the fact of logarithmic response of the human visual system [10]–[13] in measuring "contrast" between regions in an image. The present work attempts to bring this factor into consideration while defining a measure of contrast in addition to defining another measure called homogeneity within a region. The combination of these two measures made the algorithms effective in determining threshold levels. The contrast measure ensures three things. First, it assures the fact that transitions (change in the gray level) from i to $i+l$ and a transition from i to $i+k$, $l \neq k$, create different impressions. Second, a constant change in the gray level at different positions on the gray scale results in different impressions, and, finally, if the object and background intensities are interchanged, the contrast value remains unaltered.

Furthermore, provisions are also kept in two different ways for eliminating undesirable segments. In the first approach noninformative single gray level regions are eliminated by a separate

Manuscript received October 3, 1986; revised April 25, 1987. This correspondence was presented in part at the 1986 Conference on Systems and Signal Processes, Bangalore, India, December 11–13.

S. K. Pal is with the Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta 700 035, India.

N. R. Pal is with the Computer Science Unit, Indian Statistical Institute, Calcutta 700 035 India.

IEEE Log Number 8715989.

merging algorithm. The second approach, on the other hand, inherently attempts to eliminate such smaller regions irrespective of their intensity width.

The effectiveness of the algorithms, along with a comparison with three other methods [7]–[9], has been demonstrated on a set of images. A Digital computer EC-1033 has been used for the analysis.

II. Cooccurrence Matrix and Some Measures for Segmentation

A. Cooccurrence Matrix

Let $F=[f(x,y)]$ be an image of size $P\times Q$, where $f(x,y)$ is the gray value at (x,y) and $f(x,y)\in G_L=\{0,1,2,\cdots,L-1\}$, the set of gray levels. The cooccurrence matrix (or the transition matrix) of the image F is an $L\times L$ dimensional matrix that gives an idea about the transition of intensity between adjacent pixels. In other words, the (i,j)th entry of the matrix gives the number of times the gray level j follows the gray level i (i.e., the gray level j is an adjacent neighbor of the level i) in a specific fashion.

Let a denote the (i,j)th pixel in F and b denote one of the eight neighboring pixels of a, i.e.,

$$b\in a_8=\{(i,j-1),(i,j+1),(i+1,j),(i-1,j),(i-1,j-1),(i-1,j+1),(i+1,j-1),(i+1,j+1)\}.$$

Define

$$t_{lk}=\sum_{\substack{a\in F\\ b\in a_8}}\delta$$

where $\delta=1$ if the gray level value of a is l and that of b is k; $\delta=0$, otherwise.

Obviously, t_{lk} gives the number of times the gray level k follows gray level l in any one of the eight directions. The matrix $T=[t_{lk}]_{L\times L}$ is, therefore, the cooccurrence matrix of the image F. One may get different definitions of the cooccurrence matrix by considering different subsets of a_8, i.e., considering $b\in a_8'$ where $a_8'\subseteq a_8$.

The cooccurrence matrices may be either nonsymmetric or symmetric. One of the nonsymmetrical forms can be defined considering

$$t_{lk}=\sum_{i=1}^{P}\sum_{j=1}^{Q}\delta \tag{1}$$

with

$\delta=1$, if $f(i,j)=l$ and $f(i,j+1)=k$ or $f(i,j)=l$ and $f(i+1,j)=k$;

$\delta=0$, otherwise.

On the other hand, the following definition of t_{lk} gives a symmetric cooccurrence matrix:

$$t_{lk}=\sum_{i=1}^{P}\sum_{j=1}^{Q}\delta$$

where

$\delta=1$, if $f(i,j)=l$ and $f(i,j+1)=k$ or $f(i,j)=l$ and $f(i,j-1)=k$ or $f(i,j)=l$ and $f(i+1,j)=k$ or $f(i,j)=l$ and $f(i-1,j)=k$; (2)

$\delta=0$, otherwise.

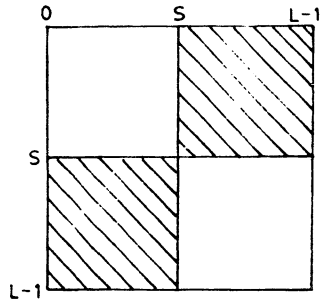


Fig. 1. Pictorial representation of busyness measure.

B. Measures for Thresholding

Since the cooccurrence matrix contains information regarding the spatial distribution of gray levels in the image, several workers have used them for segmentation. For thresholding at gray level s, Weszka and Rosenfeld [7] defined the busyness measure as follows:

$$\text{busy}(s)=\sum_{i=0}^{s}\sum_{j=s+1}^{L-1}t_{ij}+\sum_{i=s+1}^{L-1}\sum_{j=0}^{s}t_{ij}. \tag{3}$$

The cooccurrence matrix used in (3) is symmetric (using (2)).

The sum of the entries of the shaded portion in Fig. 1 represents the busyness measure for the level s. For an image with only two regions, say, object and background, the value of s for which the minimum of busy(s) occurs gives the threshold. Similarly, for an image having more than two regions, the busyness measure provides a set of minima corresponding to different thresholds.

Deravi and Pal [8] have given a measure for the conditional probability of transition from one region to another using a nonsymmetric cooccurrence matrix:

$$t_{lk}=\sum_{i=1}^{P}\sum_{j=1}^{Q}\delta \tag{4}$$

with

$\delta=1$, if $f(i,j)=l$, $f(i,j+1)=k$, and $f(i+1,j)=k$;

$\delta=0$, otherwise.

If the threshold is at s, the conditional probability of the intensity transition from the region $[0,s]$ to $[s+1,L-1]$, i.e., the probability of any intensity level from the class $[0,s]$ being followed by a level from the class $[s+1,L-1]$ in the fashion given by (4), is

$$P_1=\frac{\sum_{i=0}^{s}\sum_{j=s+1}^{L-1}t_{ij}}{\sum_{i=0}^{s}\sum_{j=0}^{s}t_{ij}+\sum_{i=0}^{s}\sum_{j=s+1}^{L-1}t_{ij}}. \tag{5a}$$

Similarly, the conditional probability of the intensity transition from the region $[s+1,L-1]$ to $[0,s]$ is

$$P_2=\frac{\sum_{i=s+1}^{L-1}\sum_{j=0}^{s}t_{ij}}{\sum_{i=s+1}^{L-1}\sum_{j=s+1}^{L-1}t_{ij}+\sum_{i=s+1}^{L-1}\sum_{j=0}^{s}t_{ij}}. \tag{5b}$$

The conditional probability $P_c(s)$ of transition across the boundary is defined as

$$P_c(s)=(P_1+P_2)/2. \tag{6}$$

The lower the value of $P_c(s)$, the lower is the probability of transition across the boundary between two classes $[0, s]$ and $[s+1, L-1]$. That means, a minimum of $P_c(s)$ will correspond to a threshold such that most of the transitions are within the class and few are across the boundary. Therefore, a set of minima of $P_c(s)$ would be obtained corresponding to different thresholds in F.

Chanda *et al.* [9] have also used the cooccurrence matrix for thresholding. They defined an average contrast (AVC) measure as

$$\mathrm{AVC}(s) = \frac{\sum_{i=0}^{s}\sum_{j=s+1}^{L-1} t_{ij}*(i-j)^2}{\sum_{i=0}^{s}\sum_{j=s+1}^{L-1} t_{ij}} + \frac{\sum_{i=s+1}^{L-1}\sum_{j=0}^{s} t_{ij}*(i-j)^2}{\sum_{i=s+1}^{L-1}\sum_{j=0}^{s} t_{ij}}. \tag{7}$$

AVC(s) shows a set of maxima corresponding to the thresholds among various regions in F.

In the computation of t_{ij} they considered only vertical transitions in the downward direction. Note here that all of the three measures described before are basically based on some weighted combinations of the number of entries in the shaded and blank regions of Fig. 1.

III. Segmentation Based on Contrast and Homogeneity Measures

In this section we present an algorithm for segmentation on the basis of the information of contrast and homogeneity (between/within the regions in F) as obtained from the cooccurrence matrix (1). The concept of the human visual system has been incorporated in the contrast measure to make the segmentation closer to the way a human being makes use of the intensity distribution for segmentation (of course, human beings use high level knowledge, too). Such a segmented image, when used for enhancement, will result in a better enhanced image as the method of segmentation tries to simulate the way a human being perceives the intensity variation. Before describing the algorithm, let us first explain some facts of the human visual system.

A. Human Psychovisual Facts

In psychophysiology contrast C refers to the ratio of difference in luminance of an object B_0 and its immediate surrounding B [12], i.e.,

$$C = |B_0 - B|/B = \frac{\Delta B}{B}. \tag{8}$$

The perceived grayness of a surface depends on its local background and the perceived contrast remains constant if the measure C of the contrast between object and local background remains constant.

The visual increment threshold (or just noticeable difference) is defined as the amount of light ΔB_T necessary to add to a visual field of intensity B such that it can be discriminated from a reference field of same intensity B. It, therefore, gives a limit for a perceivable change in luminance or intensity.

At the low-intensity near absolute visual threshold (the mere presence or absence of light intensity detectable under dark-adapted condition), the visual increment threshold ΔB_T is constant. With increasing B, ΔB_T can be approximated by a linear function of B, i.e., $\Delta B_T = \alpha \cdot B$ (the factor α is called the Weber ratio). When this linear relationship is valid, one speaks of "Weber behavior."

Fig. 2 presents such a characteristic response in the $\log \Delta B_T - \log B$ plane. The Weber behavior is characterized by unit slope of the curve. The preceding region with slope 1/2 is known as the De Vries-Rose region, characterized by $\Delta B_T = \alpha_1 \sqrt{B}$, which is a valid approximation only followed in a small restricted region. Under saturation, the visual increment threshold ΔB_T does not follow the Weber behavior, and this deviation from Weber behavior is shown by the dashed line [10].

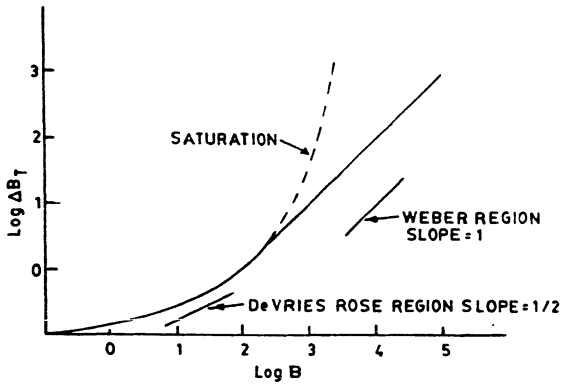


Fig. 2. Variation of $\log \Delta B_T$ with $\log B$ (in arbitrary scale) [10].

Therefore, if the brightness value of an object is higher (lower) than its surrounding or background or a reference intensity B by such an amount that it corresponds to a point on or above the curve (Fig. 2), the object will only then appear brighter (darker), i.e., discriminable to the human visual system (HVS). Furthermore, an equal amount of ΔB value created at a different background intensity (B value) does not result in an equal perceivable change to the HVS. For example, the discrimination ability in the De Vries-Rose region is greater than in the Weber region, and this ability decreases with an increase in the value of B. The possible reason for this deterioration in discrimination ability can be attributed to nonlinearities inherent in the visual system.

B. Measures of Contrast and Homogeneity

It has already been discussed that the problem of segmentation is to partition the set G_L of gray levels into some nonintersecting subsets such that each segment is as homogeneous as possible while the contrast between any segment and its neighboring segments is as high as possible. Two such measures, namely, the contrast of a segment with its neighboring segments and the homogeneity of a segment, are defined in the following. A composite measure of the two is then used to select the threshold levels in the image F.

Let the gray levels ranging from K to M form one of the segments, say R_1, of the image F, i.e., $R_1 = [K, M], K \leq M$. Define $C^B_{K,M}$, the contrast of the segment R_1 with respect to other segments, as follows:

$$C^B_{K,M} = \frac{\sum_{i\in R_1}\sum_{j\notin R_1} t_{ij}*W_{ij}}{C_1 * \sum_{i\in R_1}\sum_{j\notin R_1} t_{ij}} = \frac{\sum_{i=K}^{M}\sum_{\substack{j<K \\ \text{or} \\ j>M}} t_{ij}*W_{ij}}{C_1 * \sum_{i=K}^{M}\sum_{\substack{j<K \\ \text{or} \\ j>M}} t_{ij}}. \tag{9}$$

From Section III-A and Fig. 2 it is seen that because of the logarithmic behavior of the HVS the discrimination ability decreases with an increase in the value of background intensity B. This is what is also reflected by the contrast measure C defined in (8). To incorporate this property, we have introduced the weighting factor W_{ij} in (9).

However, note that the direct use of $\Delta B/B$ for W_{ij} will obviously change the contrast value if the object and background intensities are interchanged. Since this is not intuitively appeal-

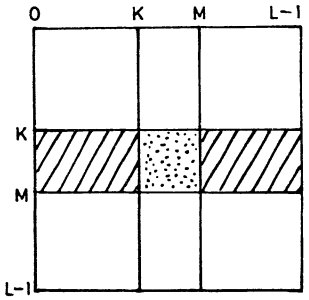


Fig. 3. Pictorial representation of contrast and homogeneity measures.

ing, W_{ij} may be defined as either

$$W_{ij} = |i-j|/(i+j), \tag{10a}$$

$$W_{ij} = |i-j|/\max\{i,j\}, \tag{10b}$$

or

$$W_{ij} = |i-j|/\min\{i,j\}. \tag{10c}$$

It is, therefore, seen that W_{ij} ensures equal contribution to (9) when the object and background intensities are interchanged in addition to the property of deterioration of discrimination ability with increase in the background intensity.

In case of a complex image, the visual system does not adapt to a single intensity level; instead, it adapts to an average level which depends on the nature of the image [14]. It is, therefore, more logical to choose (10a) for W_{ij}. In the denominator of (9) the term $\Sigma\Sigma t_{ij}$ is used to make the measure independent of the size of regions while the constant C_1 is introduced to make $0 \leqslant C^B_{K,M} \leqslant 1$. C_1 obviously depends on the choice of W_{ij} and is equal to the maximum possible value of W_{ij}. Therefore,

$$C_1 = \begin{cases} (L-1)/(L+1), & \text{for (10a)} \\ (L-1)/L, & \text{for (10b)} \\ (L-1), & \text{for (10c)} \end{cases}$$

where L is the maximum level in F.

Therefore, it appears from (9) that if $i=j$, then $C^B_{K,M}=0$, i.e., the contrast in $[K,M]$ is minimum. On the other hand, if $i=1$ and $j=L$, then the contrast is maximum ($=1$) since for all t_{ij} that are considered in $C^B_{K,M}$ the values of W_{ij} become equal to C_1, thereby making the numerator and denominator same.

The t_{ij}s considered in (9) are shown by the shaded portion in Fig. 3, which gives the total number of transitions across the boundary of the segment R_1, i.e., from the region $[K,M]$ to its outside.

Again, define $C^W_{K,M}$, the homogeneity of the region $[K,M]$, as

$$C^W_{K,M} = 1 - \frac{\sum_{i\in R_1}\sum_{j\in R_1} t_{ij} * |i-j|}{(L-1) * \sum_{i\in R_1}\sum_{j\in R_1} t_{ij}} = 1 - \frac{\sum_{i=K}^{M}\sum_{j=K}^{M} t_{ij} * |i-j|}{(L-1) * \sum_{i=K}^{M}\sum_{j=K}^{M} t_{ij}}. \tag{11}$$

The t_{ij}s considered in (11) are shown by the dotted portion in Fig. 3. $C^W_{K,M}$ lies in the interval [0,1], i.e, $0 \leqslant C^W_{K,M} \leqslant 1$. If R_1 is perfectly homogenous, $K=M$ (i.e., the region contains only one gray level), then $C^W_{K,M}=1$ as $|i-j|=0$ for all i and j.

With the decrease of the homogeneity of R_1, $|i-j|$ increases and approaches $(L-1)$. As a result, the ratio in (11) approaches to unity and $C^W_{K,M}$ tends to zero. Thus $C^B_{K,M}$ and $C^W_{K,M}$ are found to increase with increase in contrast and homogeneity, respectively, in a region $[K,M]$. On the basis of these two measures we define a composite measure:

$$g_{K,M}(C^W_{K,M}, C^B_{K,M}) = C^W_{K,M} * C^B_{K,M}. \tag{12}$$

Since $0 \leqslant C^W_{K,M} \leqslant 1$ and $0 \leqslant C^B_{K,M} \leqslant 1$, g also lies between 0 and 1. The level at which g attains a maximum value can, therefore, be considered a boundary (or threshold) between regions.

C. Extraction of Starting Regions — Algorithm 1

To extract the thresholds in an image F, we start with $R_1 = [0,0]$ and increase the size of R_1 one by one to the right side of the gray scale until we get a local maximum of $g_{K,M}$, i.e., the process is started with $K=0$, $M=0$, and M is incremented one by one until $g_{K,M}$ attains a maximum value. If the maximum occurs at the gray level K_1, then K_1 corresponds to a threshold and gray levels ranging from 0 to K_1 represent a region of the image F. Then we start with $K=M=K_1+1$, and the process is repeated as described until we get the next maximum, say at K_2. The gray levels ranging from K_1+1 to K_2 thus constitute another segment. In this way the process is carried on until the entire gray scale is exhausted.

Merging of Single-Valued Regions: The composite measure $g_{K,M}$ is found to be very sensitive to highly uniform regions. In other words, a region containing only one gray level is likely to be detected as a separate segment, although one does not desire to have such a segment. To avoid this we have suggested here a merging algorithm whereby such a single-valued segment is accepted if the transitions within the segment are higher than those across it. This criterion enables one to retain only those regions which have significant size in the spatial domain.

Therefore, the algorithm first of all finds out the regions of single gray level and determines whether such regions should be merged or not. If it decides to merge a region, then the next task is to determine the adjacent region (left or right) to which it is to be merged.

Let $R_i = [M,M]$ be the region under consideration. Let T_W be total number of transitions within the region R_i: then

$$T_W = t_{M,M}. \tag{13}$$

If T_0 is the total number of transitions from R_i to all other outside regions, then

$$T_0 = \sum_{i=M}^{M}\sum_{j\neq M} t_{ij}. \tag{14}$$

Decision Rule: If $T_W > T_0$, then the size of the region can be taken as reasonably big. Therefore, accept the region; do not merge it; otherwise, merge the region to either of its adjacent regions. Now the problem is to select the adjacent region to which it is to be merged.

Case 1: If $M=0$, merge R_i to the right adjacent region.

Case 2: If $M=L-1$, merge R_i to the left adjacent region.

Case 3: If $0<M<L-1$, we proceed as follows. Let T_L be the total number of transitions from R_i to its left adjacent region. If the left adjacent region contains gray levels ranging from L_1 to L_2, then

$$T_L = \sum_{i=M}^{M}\sum_{J=L_1}^{L_2} t_{ij}. \tag{15}$$

Suppose T_R is the total number of transitions from R_i to the right adjacent region and the right adjacent region contains gray levels ranging from R_1 to R_2, then

$$T_R = \sum_{i=M}^{M}\sum_{j=R_1}^{R_2} t_{ij}. \tag{16}$$

If $T_L > T_R$, then merge R_i to the left adjacent region; otherwise, merge it to the right adjacent region.

The foregoing decision rule can also be formulated in a more general way as follows. Let T_i be the total number of transitions from R_i to all regions including itself, i.e.,

$$T_i = \sum_{i=M}^{M} \sum_{j=0}^{L-1} t_{ij} \tag{17}$$

and

$$\theta = T_W / T_i, \qquad \theta < 1.$$

Now if $\theta > \theta_H$, accept the region, otherwise, merge it, where θ_H is some preassigned threshold value. $\theta_H = 0.5$ obviously gives the original decision rule. The advantage of defining the decision rule in this manner is that in an interactive environment one can change θ_H, if necessary, and compare the results to pick up the most appropriate value of θ_H for a particular type of image.

D. Incorporating Size of Region — Algorithm 2

In Algorithm 1 we have incorporated a separate merging phase to eliminate the undesirable small (noninformative) regions. The following algorithm introduces the concept of size of regions into the composite measure $g_{K,M}$ (12) and thereby attempts to eliminate the need of a separate merging phase.

Let us define

$$g'_{K,M} = g_{K,M} * \left(\frac{A_R}{A}\right) * \left(\frac{1}{d_R}\right)^{\alpha} \tag{18}$$

where

- A_R size of the region $[K, M]$, i.e., number of pixel intensities in F which are in the range $[K, M]$,
- A size of the image, $= P * Q$,
- α parameter, with $0 \leqslant \alpha \leqslant 1$,
- d_R length of the region $[K, M]$, $= M - K + 1$.

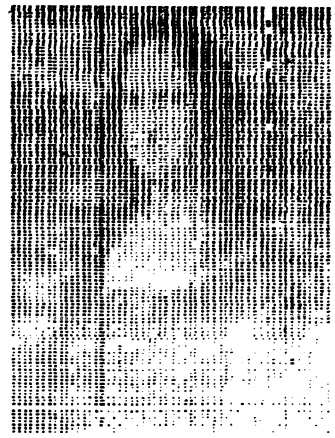
The first factor (A_R/A) gives more weight to the regions of greater size, resulting in an increase in the $g'_{K,M}$ value. Thus if the histogram of an image has a sharp valley, the $g'_{K,M}$ value is very likely to decrease when M is set at the valley point. On the other hand, for a small region with a weak valley, such a decrease in g' value (for detecting the region) is less likely to occur. In such a situation, note that the following region (valley) becomes much less likely to be detected because of the cumulative increase of region size d_R and hence A_R.

To circumvent this situation, a second factor $(1/d_R)^{\alpha}$ is introduced which compensates the foregoing effect by reducing the value of g' with an increase in d_R. α controls the rate of decrease in the g' value. As α increases, the number of regions is likely to increase. To extract the thresholds under this algorithm we proceed in exactly the same way as we have mentioned earlier for Algorithm 1, considering $g'_{K,M}$ for $g_{K,M}$.

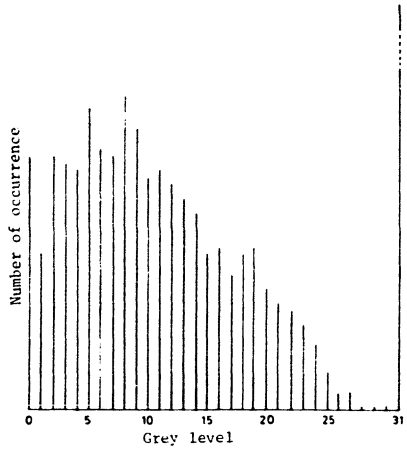
IV. Implementation and Results

The segmentation algorithms are described in the Appendix. The algorithms are implemented on a set of four different images [14] having dimension 64×64 with 32 gray levels. Figs. 4(a), 5(a), 6(a), and 7(a) represent the original input images, while Figs. 4(b), 5(b), 6(b), and 7(b) represent the corresponding gray-level histograms. These images are produced on a line printer by over printing different character combinations for different gray levels.

Fig. 4(a) represents an image of Mona Lisa. Note that the gray-level histogram (Fig. 4(b)) is almost unimodal (having some local minima). When the proposed Algorithm 1 (without merging) is applied to it, four thresholds, namely, 0, 1, 6, and 17 are produced. The corresponding segmented image is shown in Fig. 4(g), where different segments are represented by different textures. When the merging algorithm is applied to it, the segment [1, 1] (Table I) is merged to its right adjacent segment. The segmented image so obtained after merging is shown in Fig. 4(c).



(a)



(b)

Fig. 4. (a) Input image of Mona Lisa. (b) Histogram. (c) Segmented image by proposed method 1. (d) Segmented image by [7]. (e) Segmented image by [8]. (f) Segmented image by [9]. (g) Segmented image by proposed method 1 (before merging). (h) Segmented image by proposed method 2.

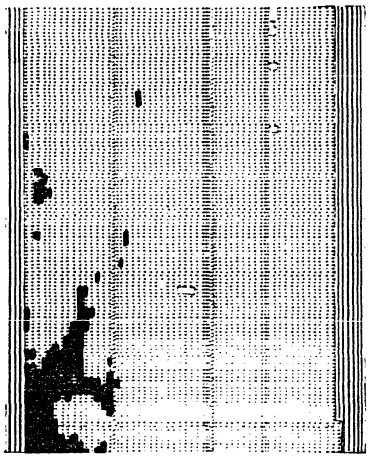
Comparing Fig. 4(c) and Fig. 4(g), we find that an undesirable region exists inside the hair of Mona Lisa, which after being merged results in a more meaningful segmentation (Fig. 4(c)). Fig. 4(g) is shown, as an illustration, only to demonstrate the effect of the merging algorithm in selecting final thresholds.

Fig. 5(a) is an image of Abraham Lincoln, and the corresponding gray-level histogram (Fig. 5(b)) is found to have a number of deep valleys. The thresholds (before and after merging) generated by the proposed method 1 are shown in Table I. The output segmented image is shown in Fig. 5(c).

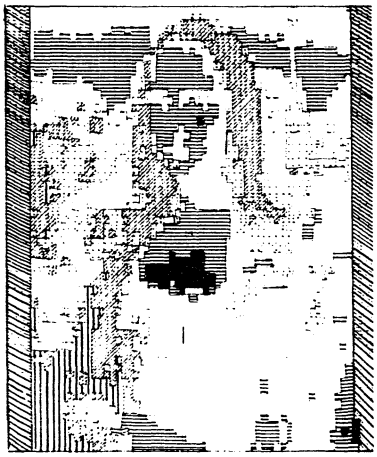
To demonstrate the validity of the method 1 for images with flat and wide valleys in their histogram, the algorithm is applied to the image of a jet (Fig. 6(a)). One can see in Fig. 6(a) that the right wing of the jet has vanished inside the cloud in such a way that apparently it is very difficult to trace the boundary of the right wing. Algorithm 1 is found to be successful in separating out that wing from the cloud. Fig. 6(c) represents the segmented image, while the thresholds are shown in Table I. As a typical illustration, Fig. 6(i) shows the plot of the $g_{K,M}$ value for the image of the jet.

Fig 7(a) represents the image of a biplane having two dominant modes in its histogram (Fig. 7(b)). From Fig. 7(c) the object is found to be clearly separated from the background.

(c)



(d)



(e)

(f)

(g)

(h)

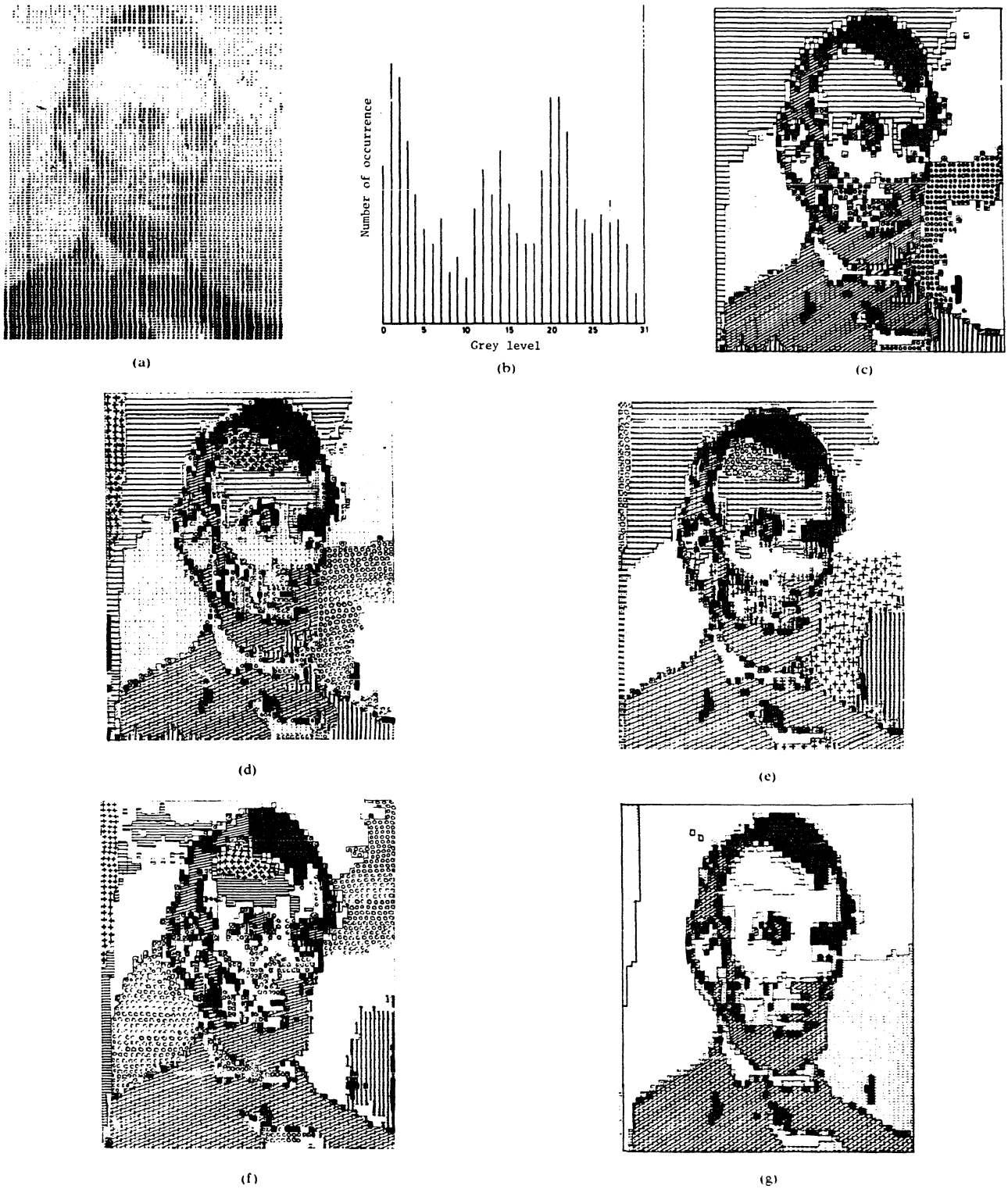
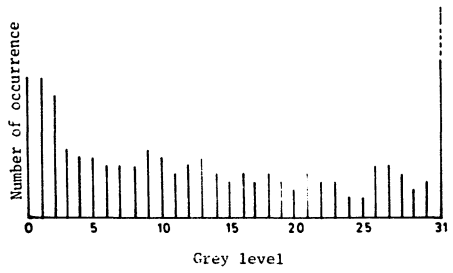
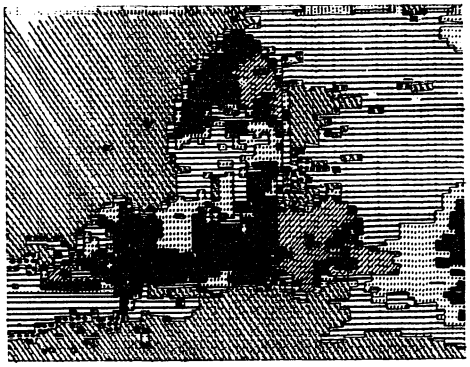


Fig. 5. (a) Input image of Lincoln. (b) Histogram. (c) Segmented image by proposed method 1. (d) Segmented image by [7]. (e) Segmented image by [8]. (f) Segmented image by [9]. (g) Segmented image by proposed method 2.

(a)

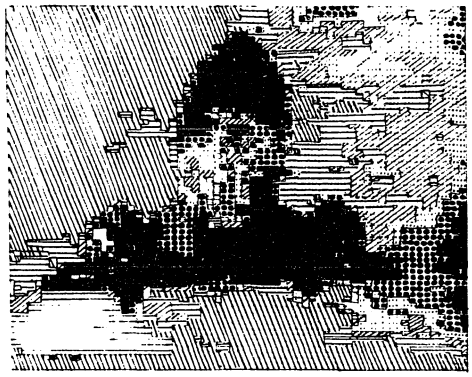


(b)

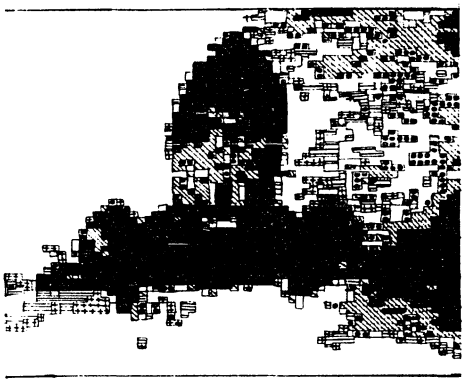


(c)

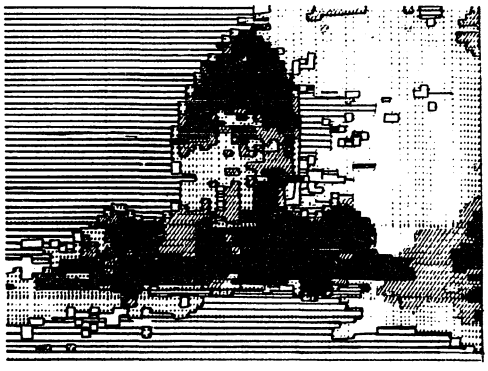
(d)



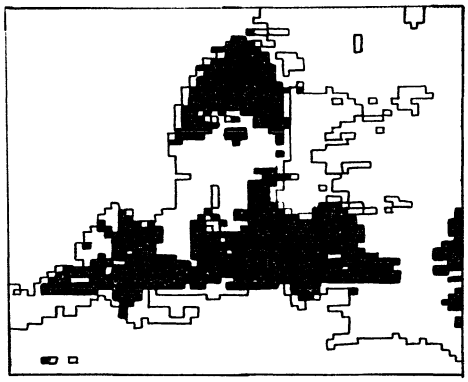
(e)



(f)



(g)



(h)

Fig. 6. (a) Input image of jet. (b) Histogram. (c) Segmented image by proposed method 1. (d) Segmented image by [7]. (e) Segmented image by [8]. (f) Segmented image nby [9]. (g) Segmented image by proposed method 1 with $\theta = 0.75$. (h) Segmented image by proposed method 2. (i) Plot of $g_{K,M}$. (j) Plot of $g'_{K,M}$ for $\alpha = 0.6$.

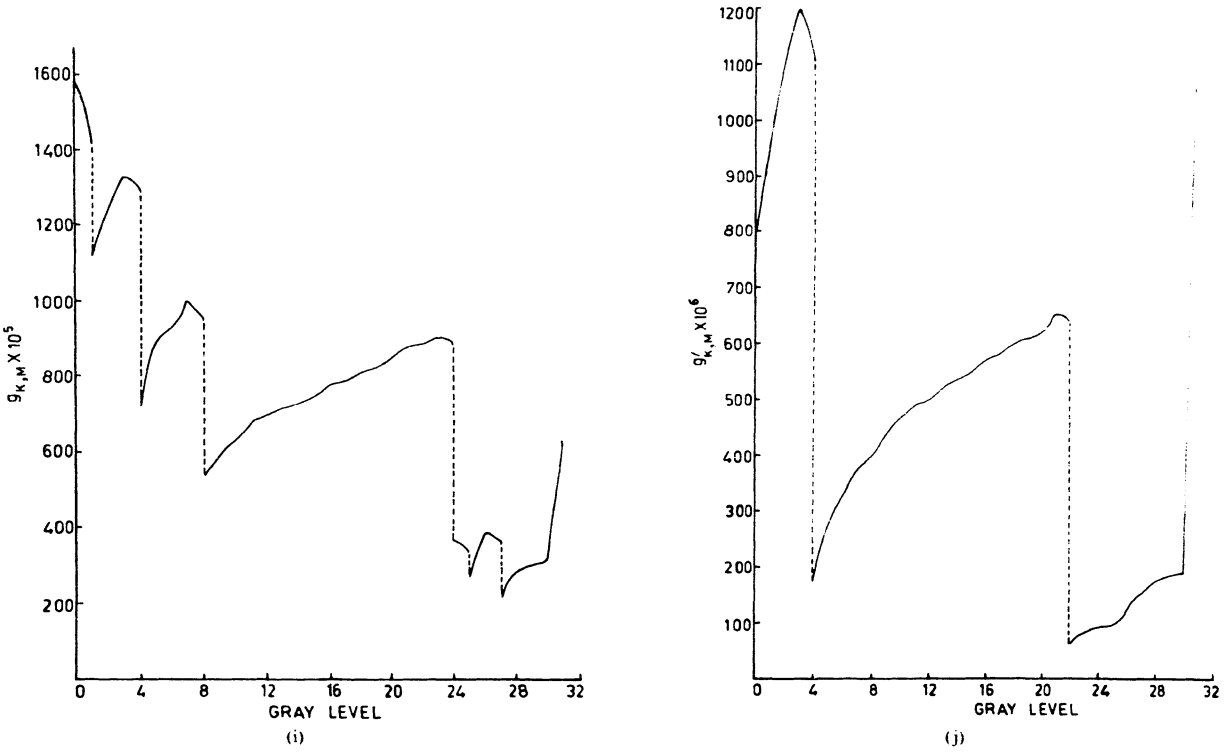


Fig. 6. *Continued.*

TABLE I
THRESHOLDS FOR VARIOUS METHODS

Image	Proposed Method: Before Merging	Proposed Method: After Merging	Method of Wezska and Rosenfeld [7]	Method of Deravi and Pal [8]	Method of Chanda *et al.* [9]
Mona Lisa Fig. 4	0,1,6,17	0,6,17	0,30	0,3,6,17, 25,30	0,3,6,17, 28,30
Lincoln Fig. 5	0,4,9,13,17, 18,24	0,4,9,13, 17,24	0,4,9,12,17, 24,30	4,9,12,17, 23,30	5,10,12,17, 22,27,30
Jet Fig. 6	0,3,7,23, 24,26	0,3,7,24, 26	0,3,5,7,18, 29	3,7,11,18 29	7,11,13,18, 21,23
Biplane Fig. 7	0,1,2,10 17,18,23	1,10,17 23	0,10,17, 22,30	0,2,10,17, 22,30	0,10,14,17, 21,29

TABLE II
THRESHOLDS FOR DIFFERENT α

Image	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 1.0$
Mona Lisa Fig. 4	17,28	6,25	0,6,25,28	0,6,17,24, 28	0,6,16,24, 25,28	0,3,6,9,14, 22,24,25,28
Lincoln Fig. 5	5	4,17	4,9,17	4,9,17,29	4,9,17,24, 29	2,4,7,9,15, 17,23,29
Jet Fig. 6	7	3,23	3,21	3,21	3,7,16,23, 28	0,2,3,7,11, 13,16,19, 23,24,28
Biplane Fig. 7	10	10	8,10	8,10	0,7,10	2,7,9, 23,29

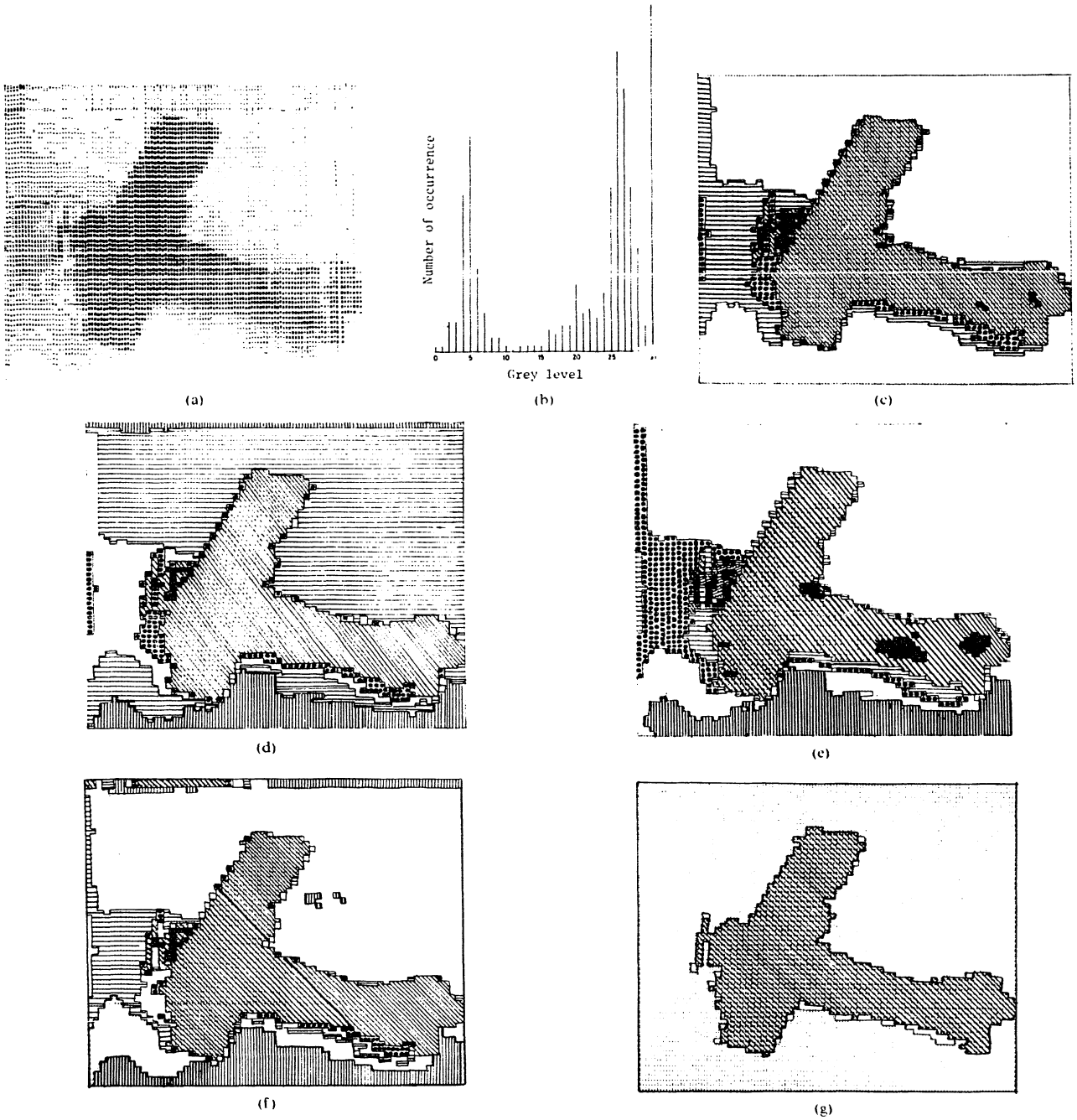


Fig. 7. (a) Input image of biplane. (b) Histogram. (c) Segmented image by proposed method 1. (d) Segmented image by [7]. (e) Segmented image by [8]. (f) Segmented image by [9]. (g) Segmented image by proposed method 2.

The aforementioned results were obtained by using (10a) while computing W_{ij} of the contrast measure. Experiments were also carried out with (10b), and the corresponding performances were found to be almost similar.

Table II shows the thresholds for these images obtained by the proposed Algorithm 2 for different values of α. As discussed in Section III-D, the number of segments is found to increase with increases in the value of α. Note, too, from Table II that the level 17 for Mona Lisa (unimodal histogram), which was detected at $\alpha = 0.2$, was found to be missing for values of $\alpha = 0.4$ and 0.6. The threshold at 17 or its around again appeared for $\alpha \geq 0.7$. The case for jet is similar, having an almost flat and wide valleyed histogram where the level 7, which was detected at $\alpha = 0.2$, was lost at $\alpha = 0.4$, 0.6, and 0.7 and reappeared for higher α values. Such a situation did not arise for the Lincoln and biplane images, which have deep valleys. This fact can be explained as follows.

Suppose the first threshold is detected at a gray level $l = l_1$ for $\alpha = \alpha_1$, which means that g' kept on increasing up to gray level l_1 and at $(l_1 + 1)$, g' started falling. Now if for a higher value of $\alpha = \alpha_2$, $\alpha_2 > \alpha_1$, the first threshold is detected earlier at a gray

level $l = l_2$, $l_2 < l_1$, then we can say that l_2 is detected, as $(1/d_R)^\alpha$ has reduced the value of g' at $(l_2 + 1)$ to a value which is less than the value of g' at l_2. The threshold at $l = l_1$, which was detected with $\alpha = \alpha_1$, may not now be detected with an α not much higher than α_1 because no sharp fall will occur in the g value as the histogram has weak valleys. Moreover, the value of $(1/d_R)^\alpha$ with $d_R = l_1 + 1$ and $\alpha = \alpha_1$ may be smaller than $(1/d_R)^\alpha$ with $d_R = l_1 - l_2$, and $\alpha = \alpha_2$ as d_R in the second case is reduced to $l_1 - l_2$ from $l_1 + 1$, unless α_2 is very high. Therefore, unless α_2 is sufficiently high, the threshold at l_1 will not be detected. However, if the histogram has a deep valley at l_1, it will be detected because the fall in the g value will be sharp.

Figs. 4(h), 5(g), 6(h), and 7(g) show the segmented images for Algorithm 2 when α is considered to be 0.7, as an illustration. Fig. 6(j) shows the plot of the $g'_{K,M}$ value for the image of the jet.

Comparison with the Existing Algorithms

To compare the performance of the algorithms with those of some of the existing algorithms based on the cooccurrence matrix, we have considered algorithms of Weszka and Rosenfeld (3) [7], Deravi and Pal (4)–(6) [8], and Chanda *et al.* (7) [9]. The thresholds obtained by these methods are also shown in Table I.

Equation (3) is found to fail to extract all the meaningful regions of the image of Mona Lisa. It has selected only three segments (Fig. 4(d)) with thresholds at 0 and 30; as a result, most of the important information is lost.

Equations (4)–(6), on the other hand, detected two extra segments in the chest and one extra region in the hair (Fig. 4(e)), while (7) also produced two similar segments (one of them is of smaller size than the corresponding one produced by (4)–(6) (Fig. 4(f)). From Fig. 4(c) it is seen that the regions generated by the proposed method 1, where these additional regions are absent, create a better impression to the eye. However, for Algorithm 2 (for $\alpha = 0.7$) the hair of Mona Lisa did not split, and it has two extra regions in her chest (Fig. 4(h)).

For the image of Lincoln (Fig. 5) all the methods except the present ones have divided the forehead into two regions. Furthermore, Algorithm 1 and the method by Weszka and Rosenfeld have divided the beard into two regions, which is not the case for the other three methods. Finally, Algorithm 2 created the smallest number of segmented regions (Fig. 5(g)) and created a better impression the eye.

In the case of the jet, (7) failed to discriminate between the cloud and the right wing (Fig. 6(f)) while the other two methods, like ours, are successful in doing so (Fig. 6(c)–(e)). Algorithm 1 and (3) produced comparable results, while the result produced by (4)–(6) (Fig. 6(e)) seems to create a better impression to the eye, but if we alter the value of θ_H from 0.5 to 0.75 in the merging algorithm, the first two regions (Table I) are merged and the resulting image (Fig. 6(g)) is found to be much improved. However, Fig. 6(h) (by Algorithm 2), which has least number of segments, created the best result.

In the case of the biplane all the methods were able to detect its contour (Fig. 7). However, (4)–(6) are found to generate two additional regions inside the tail of the biplane which are absent for other cases. The background is found to be clustered in two parts by all but our methods. Furthermore, all the methods except ours divided the shade of the biplane into three or more regions, which is two and one for Algorithms 1 and 2, respectively.

From Table I it appears that (3), (4)–(6), and (7) detected, except for the jet, a threshold at the end of the grey scale (e.g., 30 for Mona Lisa, Lincoln, and the biplane). These thresholds correspond to some undesirable regions at the frame of the images. The incorporation of the factor W_{ij} in (9), which accounts for the nonlinear behavior of HVS, has been found to be able to eliminate such occurrences. However, Algorithm 2 failed to eliminate such segments for higher value of α.

V. Conclusion and Discussion

Two algorithms for image segmentation are described using the measures of homogeneity and contrast within/between regions of an image. The first algorithm has a separate merging phase to eliminate the undesirable segments. The second algorithm, on the other hand, eliminated the need of such a separate merging phase by considering the region size while extracting the thresholds. Both of the algorithms simulate the way in which human beings perceive brightness variation in an image.

Note here that when a human being attempts to segment an image, in addition to his logarithmic response to contrast he makes use of certain other higher level knowledge which is not considered in the proposed algorithms. Hence it is not expected that the proposed algorithms would segment an image exactly in the same way a human observer does. However, the incorporation of the weighting function W_{ij} to simulate only the contrast response of the human visual system has resulted in consistently better segments for images with different types of histograms.

It is found that none of the existing methods could generate as consistently good segments as those of the proposed algorithms for all images. Some of the existing methods generated extra segments for some images, while some methods have failed to detect meaningful segments. Both of the proposed algorithms are suitable for an interactive environment.

It is seen that the choice of α in Algorithm 2 is critical. As α increases, the number of regions also increases, resulting in an acceptable segmentation around 0.5. However, for a good image having sharp valleys in the gray level histogram, the typical value of α may be less than 0.5. On the other hand, an α greater than 0.5 can be taken for images having weak valleys in the histogram or a unimodal histogram.

The proposed methods are expected to be less sensitive to noise as the psychovisual facts have been considered. The work can be extended to three-dimensional images.

Appendix

Algorithm 1: Extraction of Initial Thresholds

Step 1 $K = 0$, $M = 0$, *number of thresholds* = 0.
Step 2 Compute *previous value* = $g_{K,M}$.
Step 3 $M = M + 1$.
If M is greater than $L - 1$, then go to step 6; otherwise, compute *current value* = $g_{K,M}$.
Step 4 If *current value* is not less than *previous value*, then *previous value* = *current value* and go to Step 3.
Step 5 *Number of thresholds* = *number of thresholds* + 1, THRESHOLD (*number of thresholds*) = $M - 1$ (*threshold* is an array that stores the threshold levels), $K = M$, go to Step 2.
Step 6 End of extraction of thresholds.

Merging Algorithm (Output of Algorithm 1 is Input to This Algorithm)

Step 1 Find the next region of single gray level. If there is none, then stop.
Step 2 (Suppose the region under consideration is $R_i = [M, M]$.) Compute T_{ii} = the number of transitions within R_i. Compute T_i = the total number of transitions from R_i to all other regions including itself. Compute $\theta = T_{ii}/T_i$.
Step 3 If θ is greater than θ_H (a predetermined value), then go to Step 1.
Step 4 If M is the first gray level with nonzero frequency, then merge the region R_i to its right adjacent region and go to Step 1.
Step 5 If M is the last gray level with nonzero frequency, then merge the region R_i to its left adjacent region and go to Step 1.

Step 6 Compute T_R and T_l, the number of transitions from R to its right and left adjacent regions, respectively.
Step 7 If T_l is greater than T_R, then merge R_i to its left adjacent region. Otherwise, merge R_i to its right adjacent region.
Step 8 Go to Step 1.

Algorithm 2

Replace $g_{K,M}$ by $g'_{K,M}$ in Algorithm 1. (Note that Algorithm 2 does not need the merging algorithm of Algorithm 1.)

ACKNOWLEDGMENT

The authors gratefully acknowledge the help of Mrs. S. De Bhowmick in typing the manuscript and Professor D. Dutta Majumder for his interest in this work.

REFERENCES

[1] T. Pavlidis, *Structural Pattern Recognition*. New York: Springer, 1977.
[2] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. New York: Academic, 1982.
[3] J. S. Weszka, "Survey of threshold selection techniques," *Comput. Graphics Image Processing*, vol. 7, pp. 259–265, 1978.
[4] A. Rosenfeld and L. S. Davis, "Iterative histogram modification," *IEEE Trans. Syst., Man, Cybern.* vol. SMC-8, no. 4, pp. 300–302, 1978.
[5] S. Peleg, "Iterative histogram modification," *IEEE Trans. Syst. Man, Cybern.*, vol. SMC-8, no. 7, pp. 555–556, 1978.
[6] S. K. Pal, R. A. King, and A. A. Hashim, "Automatic grey level thresholding through index of fuzziness and entropy," *Patt. Recog. Lett.*, vol. 1, pp. 141–146, 1983.
[7] J. S. Weszka and A. Rosenfeld, "Threshold evaluation techniques," *IEEE Trans Syst. Man, Cybern.*, vol. SMC-8, pp. 622–629, 1978.
[8] F. Deravi and S. K. Pal, "Graylevel thresholding using second-order statistics," *Patt. Recog. Lett.*, vol. 1, no. 5, pp. 417–422, 1983.
[9] B. Chanda, B. B. Chaudhuri, and D. Dutta Majumder, "On image enhancement and threshold selection using the graylevel co-occurrence matrix," *Patt. Recog. Lett.*, vol. 3, no. 4, pp. 243–251, 1985.
[10] G. Buchsbaum, "An analytical derivation of visual nonlinearity," *IEEE Trans. Biomed. Eng.*, vol. BME-27, pp. 237–242, 1980.
[11] P. Zuidema *et al.* "A mechanistic approach to threshold behavior of visual systems," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, pp. 923–934, 1983.
[12] E. L. Hall, *Computer Image Processing and Recognition*. New York: Academic, 1979.
[13] M. K. Kundu and S. K. Pal, "Thresholding for edge detection using human psychovisual phenomena," *Patt. Recog. Lett.*, vol. 4, no. 6, pp. 433–441, 1986.
[14] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1977.

Learning Behaviors of the Hierarchical Structure Stochastic Automata under the Nonstationary Multiteacher Environments and Their Applications to Intelligent Robot Manipulators

NORIO BABA, MEMBER, IEEE

***Abstract*—Learning behaviors of the hierarchical structure stochastic automata operating in a nonstationary multiteacher environment are considered. It is shown that an extended form of the algorithm proposed by Thathachar and Ramakrishnan ensures absolute expediency under some condition. As a practical application of the hierarchical structure stochastic automata, an intelligent behavior of robot manipulators going through a maze having a large number of gates that close with unknown rejecting probabilities is also considered. It is shown that the hierarchical structure stochastic automata can be successfully utilized to let robot manipulators gradually find the best way in the maze.**

Manuscript received January 25, 1987; revised May 18, 1987.

The author is with Information Science and System Engineering, Faculty of Engineering, Tokushima University, Josanjima, Tokushima City, 770 Japan.

IEEE Log Number 8715988.

I. INTRODUCTION

The study of learning automata operating in an unknown random environment was started by Tsetlin [1]. He considered the learning behaviors of finite deterministic automata under the stationary random environment $R(C_1,\cdots,C_r)$ and showed that they are asymptotically optimal under some conditions.

Following his pioneering work, Varshavskii and Vorontsova [2] found that stochastic automata also have learning performances. Since then, the learning behaviors of stochastic automata have been studied quite extensively by many researchers. The survey papers [5], [11] and the books [15], [17] contain most of the recent works in this field along with the valuable comments for the future research.

Recently, the concept of the hierarchical structure automata was introduced by Thathachar and Ramakrishnan [8] and Ramakrishnan [16]. Since many living systems follow hierarchical plans and our daily decisionmaking is done in a hierarchical fashion, the concept of the hierarchical structure automata system would become one of the most promising tools in various application areas.

In this correspondence, we consider the learning behaviors of the hierarchical structure stochastic automata operating in the general nonstationary multiteacher environment (S model) and show that an extended form of the algorithm proposed by Thathachar and Ramakrishnan [8] can ensure absolute expediency under some restrictive condition. Further, we give an example which verifies that the learning behaviors of the hierarchical structure stochastic automata can be successfully utilized to let robot manipulators in a maze gradually find the best way.

II. LEARNING MECHANISM OF THE HIERARCHICAL STRUCTURE STOCHASTIC AUTOMATA

Fig. 1 shows the learning mechanism of the hierarchical structure stochastic automata operating in the general nonstationary multiteacher environment. The hierarchy is composed of the single automaton $\bar{A}$ in the first level, r automata $\bar{A}_1,\cdots,\bar{A}_r$ in the second level, and r^{s-1} automata $\bar{A}_{j_1 j_2 \cdots j_{s-1}}$ ($j_i=1,\cdots,r$; $i=1,\cdots,s-1$) in the sth level ($s=1,\cdots,N$). Each automaton in the hierarchy has r actions.

The operation of the learning system of the hierarchical structure stochastic automata can be described as follows. Initially, all the action probabilities are set equal. $\bar{A}$ chooses an action at time t from the action probability distribution $(p_1(t),\cdots,p_r(t))'$. Suppose that α_{j_1} ($j_1=1,\cdots,r$) is the output from $\bar{A}$ and $\beta_1^{k_1 j_1}(t,\omega)$, $k_1=1,\cdots,r_1$, where ω is a point of the probability measure space (Ω, B, P) and t denotes time, is the response from the k_1th teacher. ($\beta^{k_1 j_1}(t,\omega)$ can be an arbitrary number in the closed line segment [0,1].) Depending upon the output α_{j_1}, the responses $\beta_1^{k_1 j_1}$, $k_1=1,\cdots,r_1$, and the responses from the lower levels, the first level automaton $\bar{A}$ changes its action probability vector $P(t)=(p_1(t),\cdots,p_r(t))'$. Corresponding to the output α_{j_1} at the first level, the automaton $\bar{A}_{j_1}$ is actuated in the second level. This automaton chooses an action from the current action probability distribution $(p_{j_1 1}(t),\cdots,p_{j_1 r}(t))'$. This cycle of operation repeats from the top to the bottom.

Let $\phi_{i_1 i_2 \cdots i_s}$ be the path which is composed of a sequence of actions α_{i_1}, $\alpha_{i_1 i_2}$, $\cdots$, and $\alpha_{i_1 i_2 \cdots i_s}$ chosen by the actuated automata in the first s levels ($s=1,\cdots,N$).