On hierarchical segmentation for image compression

Sambhunath Biswas [a,*], Nikhil R. Pal [b]

[a] *Machine Intelligence Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700 035, India*

[b] *Electronics and Communication Sciences Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700 035, India*

Received 27 October 1998; received in revised form 27 July 1999

Abstract

The present paper describes a hierarchical image segmentation scheme keeping in mind its use in image compression. At each level of hierarchy, the segmentation provides a *sub-image* consisting of compact, homogeneous regions. A number of thresholds based on conditional entropy of the image guides the entire process. Small regions are merged. Objective measures based on correlation and contrast have been proposed for evaluation of the segmentation technique, and the result of the proposed algorithm has been compared with those of three different existing multi-level thresholding algorithms.

Keywords: Segmentation; Threshold; Entropy; Correlation; Contrast

1. Introduction

Segmentation plays a significant role in image processing as well as in pattern recognition. Segmented regions along with their contours may be useful for designing image compression algorithms. This representation is more useful and effective because region contours are not disconnected like edges. Reported works in this area can be found in (Kunt et al., 1985, 1987; Carlsson, 1988; Shen and Rangayyan, 1997), whereas a broad over-view of segmentation can be found in (Gonzalez and Wintz, 1977; Pavlidis, 1977; Rosenfeld and Kak, 1982).

To get compact homogeneous regions (or patches), we have developed a new segmentation scheme where we have recursively used an object/background thresholding algorithm (Pal and Bhandari, 1993). Unlike the region growing (Pavlidis, 1982) or adaptive region growing (Kunt et al., 1987) technique, it provides a number of compact regions of similar gray levels for a given threshold. We call this collection of regions for a given threshold a *sub-image*. Thus, the segmentation scheme produces a number of sub-images depending on the number of computed thresholds. A strategy for merging small regions has also been suggested. We propose some quantitative indices for objective evaluation of the segmented regions.

2. Extraction of compact homogeneous regions

The compression scheme may be thought of as based on modeling compact homogeneous regions or patches using Bezier–Bernstein polynomial function. Thus, given an image, we extract the homogeneous sub-images. This can be achieved through segmentation for which there are many approaches (Weszka, 1978; Fu and Mui, 1981; Haralick and Shapiro, 1985). For example, it can be based on pixel level decision making such as iterative pixel modification or region growing or adaptive region growing, or it can be based on multilevel thresholding. Each of these categories of algorithms, except multilevel thresholding produces one region of similar gray levels at a time and, therefore, it forces local approximation for a region. Such methods may be called local thresholding schemes as decision is made at the pixel level. It does not provide any information about other regions of similar gray values. Hence from the standpoint of compression, segmentation algorithms based on local region growing are not very attractive. On the other hand, global thresholding-based segmentation algorithms, (where the entire image is partitioned by one or a few thresholds), such as multilevel thresholding algorithms (Weszka and Rosenfeld, 1978; Deravi and Pal, 1983; Chanda et al., 1985), depend on the number of local minima in the one- or two-dimensional histogram of gray values in the image. The extraction of these minima from the histogram information sometimes may not be very reliable, because all desirable thresholds may not be reflected as deep valleys in the histogram. Also, the detection of thresholds is influenced by all pixels in the image.

Several authors (Abutaleb, 1989; Kapur et al., 1985; Pal and Pal, 1989a,b; Pun, 1980, 1981) have used entropy as the criterion for object/background classification. All methods described in (Kapur et al., 1985; Pun, 1980, 1981) use only the entropy of the histogram, while the methods in (Abutaleb, 1989; Pal and Pal, 1989a,b) use the spatial distribution of gray levels, i.e., the higher-order entropy of the image. For the set of images reported in (Pal and Bhandari, 1993), authors claimed that conditional entropy of the objects and background based on Poisson distribution produced better results compared to the methods in (Kapur et al., 1985; Pal and Pal, 1989a,b; Pun, 1980, 1981; Kitler and Illingworth, 1986). All these methods produce only an object/background (two level) partitioning of the image. In our problem such a bi-level thresholding is not adequate. We propose an algorithm for hierarchical extraction of homogeneous patches using the conditional entropy thresholding method. The conditional entropy is defined in terms of the second-order co-occurrence matrix.

(a) *Co-occurrence matrix*. Let $F = [f(x,y)]$ be an image of size $M \times N$, where $f(x,y)$ is the gray value at (x,y), $f(x,y) \in G_L = \{0,1,2,\ldots,L-1\}$, the set of gray levels. The co-occurrence 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' in a specific way.

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

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

Define

$$t_{ik} = \sum_{a \in F, b \in a_8} \delta,$$

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

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

The co-occurrence matrix may again be either asymmetric or symmetric. One of the asymmetrical forms can be defined considering

$$t_{ik} = \sum_{i=1}^{M} \sum_{j=1}^{N} \delta$$

with

$$\delta = \begin{cases} 1 & \text{if } f(i,j) = i \text{ and } f(i,j+1) = k, \\ & \text{or } f(i,j) = i \text{ and } f(i+1,j) = k, \\ 0 & \text{otherwise.} \end{cases}$$

Here only the horizontally right and vertically lower transitions are considered. The following definition of t_{ik} gives a symmetrical co-occurrence matrix:

$$t_{ik} = \sum_{i=1}^{M} \sum_{j=1}^{N} \delta,$$

where

$$\delta = \begin{cases} 1 & \text{if } f(i,j) = i \text{ and } f(i,j+1) = k \\ & \text{or } f(i,j) = i \text{ and } f(i,j-1) = k \\ & \text{or } f(i,j) = i \text{ and } f(i+1,j) = k \\ & \text{or } f(i,j) = i \text{ and } f(i-1,j) = k, \\ 0 & \text{otherwise.} \end{cases}$$

(b) *Conditional entropy of a partitioned image.* The entropy of an n-state system as defined by Shannon and Weaver (1949) is

$$H = -\sum_{i=1}^{n} p_i \ln p_i, \tag{1}$$

where $\sum_{i=1}^{n} p_i = 1$ and $0 \leqslant p_i \leqslant 1$, p_i is the probability of the ith state of the system. Such a measure is claimed to give information about the actual probability structure of the system. Some drawbacks of (1) were pointed out by Pal and Pal (1989a) and the following expression for entropy was suggested:

$$H = \sum_{i=1}^{n} p_i \exp(1 - p_i), \tag{2}$$

where $\sum_{i=1}^{n} p_i = 1$ and $0 \leqslant p_i \leqslant 1$. The term $-\ln p_i$, i.e., $\ln(1/p_i)$ in (1) or $\exp(1 - p_i)$ in (2) is called gain in information from the occurrence of the ith event. Thus, one can write,

$$H = \sum_{i=1}^{n} p_i \Delta I(p_i), \tag{3}$$

where $\Delta I(p_i) = \ln(1/p_i)$ or, $\exp(1 - p_i)$ depending on the definition used.

Considering two experiments $A(a_1, a_2, \ldots, a_m)$ and $B(b_1, b_2, \ldots, b_n)$ with respectively, m and n, possible outcomes, the conditional entropy of A given b_l has occurred in B is

$$H(A \mid b_l) = \sum_{k=1}^{m} p(a_k \mid b_l) \Delta I(p(a_k \mid b_l)), \tag{4}$$

where $p(a_k \mid b_l)$ is the conditional probability of occurrence of a_k given that b_l has occurred. We can write the entropy of A conditioned by B as

$$\begin{aligned} H(A \mid B) &= \sum_{l=1}^{n} p(b_l) H(A \mid b_l) \\ &= \sum_{l=1}^{n} \sum_{k=1}^{m} p(b_l) p(a_k \mid b_l) \Delta I(p(a_k \mid b_l)) \\ &= \sum_{l=1}^{n} \sum_{k=1}^{m} p(a_k, b_l) \Delta I(p(a_k \mid b_l)), \end{aligned} \tag{5}$$

where $p(a_k, b_l)$ is the joint probability of occurrence of (a_k, b_l).

Let $p(i \mid j)$ be the probability that a gray value i belongs to the object, given that the adjacent pixel with gray value j belongs to the background, $\sum_i p(i \mid j) = 1$. Thus, for a given threshold s, the conditional entropy of the object given the background, as defined by Pal and Bhandari (1993) (using (5)) is

$$\begin{aligned} H_s(O \mid B) &= \sum_{i \in \text{object}} \sum_{j \in \text{background}} p_0(i,j) \Delta I(p_o(i \mid j)) \\ &= \sum_{i=0}^{s} \sum_{j=s+1}^{L-1} p_o(i,j) \Delta I(p_0(i \mid j)), \end{aligned} \tag{6}$$

where

$$p_0(i,j) = \frac{t_{ij}}{\sum_{i=0}^{s} \sum_{j=s+1}^{L-1} t_{ij}} \tag{7}$$

and

$$p_0(i \mid j) = \frac{t_{ij}}{\sum_{i=0}^{s} t_{ij}}, \tag{8}$$

for $0 \leqslant i \leqslant s$ and $s+1 \leqslant j \leqslant L-1$. Here t_{ij} is the frequency of occurrence of the pair (i, j). The conditional entropy of the background given the object can similarly (using (5)) can be defined as

$$H_s(B \mid O) = \sum_{i \in \text{background}} \sum_{j \in \text{object}} p_b(i,j) \Delta I(p_b(i \mid j)), \quad (9)$$

where

$$p_b(i,j) = \frac{t_{ij}}{\sum_{i=s+1}^{L-1} \sum_{j=0}^{s} t_{ij}} \quad (10)$$

and

$$p_b(i \mid j) = \frac{t_{ij}}{\sum_{i=s+1}^{L-1} t_{ij}}, \quad (11)$$

for $s+1 \leqslant i \leqslant L-1$ and $0 \leqslant j \leqslant s$. Then the total conditional entropy of the partitioned image is

$$H_T^C = H_s(O \mid B) + H_s(B \mid O). \quad (12)$$

For an image, the conditional entropy of the object given the background provides a measure of information about the object when we know about the existence of the background. Entropy is a measure of expected gain in information or expected loss of ignorance with an associated probability distribution. Thus, $H(O \mid B)$ can also be viewed as average loss of ignorance about the object when we are told about the background. Similar interpretation is applicable to $H(B \mid O)$ also. Hence, maximization of H_T^C is expected to result in a good threshold. H_T^C can also be viewed as a measure of contrast.

Let th be the correct threshold for an object/background segmentation. Now if th is used to partition the co-occurrence matrix, entries in quadrants two and four in Fig. 1 will have low frequencies, but expected to be more or less uniformly distributed. Similarly, for one and third quadrants, frequencies also will be uniformly distributed but with high values. Because within a region, frequencies of transition from one level to another will be high. However, as far as the two-dimensional probability distribution is concerned, all cell will have more or less uniform probability mass function. Now suppose the assumed threshold s is less than th, the second quadrant will have some high frequencies which are actually transitions within the object. In addition to this, it will also have actual low frequency transitions from object to background (i.e., across the boundary).

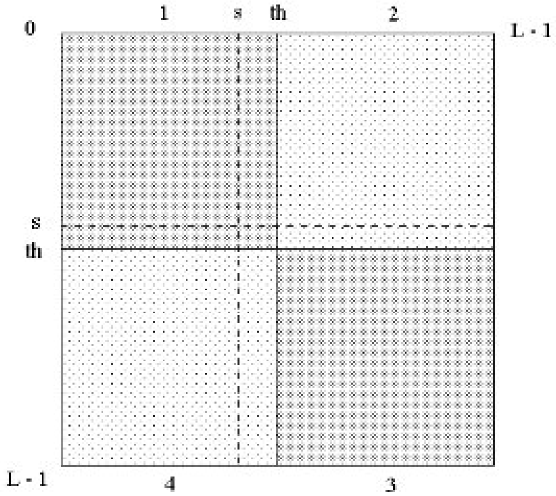

Fig. 1. Partitioning of the co-occurrence matrix for thresholding.

Thus, the quadrant two will have a highly skewed probability distribution resulting in a drastic lowering of H_T^C.

The uniformity of quadrant one will be maintained, but that of quadrants three and four will be affected causing a lowering of entropy of quadrants three and four. Similarly, if the assumed threshold is more than th, H_T^C will be reduced. Hence, its maximization with respect to s is expected to provide a good object/background segmentation.

Next, we provide a schematic description of the algorithm.

Algorithm (*Cond_threshold*(X, th)).

```
begin
   Compute Co-occurrence matrix, t = [t_ij]_{L×L}.
   s = 0; max = 0;
   th = 0; th is the threshold for segmentation
   while (s < L − 1) do
      compute H_T^C by (12)
      if (H_T^C(s) > max) then begin
                                  th = s;
                                  max = H_T^C(s)
                                  end
      s = s + 1;
   endwhile;
end;
```

In the present investigation, we use $\Delta I(p_i) = \exp(1 - p_i)$ in Eq. (12).

2.1. Partition/decomposition principle for gray images

In this section, we explore the possibility of using the object/background thresholding algorithm (*Cond_threshold*) for extraction of homogeneous patches from a gray level image for image data compression. For this purpose, we intend to partition the image into several sub-images keeping in view the following points:

- each sub-image consisting of different regions should be approximated well by some low-order function,
- number of sub-images should be as low as possible,
- homogeneity within a region and contrast between-regions should be reasonably good.

In order to achieve this goal, one can use either a multi-level thresholding algorithm (Weszka and Rosenfeld, 1978; Deravi and Pal, 1983; Chanda et al., 1985) or an object/background thresholding algorithm. The multi-level thresholding algorithm depends on the number of local minima in the histogram of the image. The extraction of these minima from the histogram information sometimes may not be very reliable, because some of them may not be strong enough to be detected by the objective function being used. The object/background algorithm, on the other hand, relies on a single threshold to extract the object from the background. We propose a scheme which repeatedly uses an object/background segmentation algorithm for extraction of homogeneous patches suitable for image compression.

Consider an L-level image $F_0(x,y)$. The input gray image $F_0(x,y)$ initially provides a threshold, s on application of the object/background thresholding algorithm. The threshold, s partitions the image $F_0(x,y)$ into two sub-images $F_{01}(x,y)$ and $F_{02}(x,y)$. The gray levels in $F_{01}(x,y)$ lies in the interval $[0,s]$ and in $F_{02}(x,y)$ it is limited to $(s, L-1]$. From the standpoint of object/background thresholding, $F_{01}(x,y)$ can be viewed as the object while $F_{02}(x,y)$ is the background without loss of generality.

To check the feasibility of global approximation of the sub-images so obtained, we approximate, first of all, $F_{01}(x,y)$ by a polynomial of order $p \leqslant q$ (q is a predefined upper limit on the order of polynomials) satisfying a criterion C. It should be noted that $F_{01}(x,y)$ may consist of a number of isolated regions or patches, say, $\Omega_1, \Omega_2, \ldots, \Omega_r$. If the approximation satisfies the criterion C then we accept the sub-image $F_{01}(x,y)$. Otherwise, even when a polynomial surface of order q cannot approximate the sub-image subject to C, we compute the variance in each of the regions. Next, we fit a global surface of order q over the entire sub-image and a local surface of order less than q over the residual errors (defined with respect to surface of order q) of the most dispersed region. This may give rise to one of the following four different situations:

(1) the criterion C is satisfied for the most dispersed region (with respect to global and local surface fitting) and also for rest of the regions (with respect to global fitting),

(2) C is satisfied for the most dispersed region but not for rest of the regions,

(3) C is not satisfied for the most dispersed region but satisfied for rest of the regions,

(4) C is not satisfied for both the most dispersed region and rest of the regions.

In situation (1), both local and global fits are satisfied. Hence, it implies that all segmented regions or surface patches are homogeneous and we accept the sub-image.

In situation (2), we additionally fit a local surface of order less than q over the residual errors (defined with respect to surface of order q) of the second most dispersed region. The process may continue for all regions in the sub-image, only in case of failure for the global surface approximation. But if the local surface fit fails to satisfy the criterion C at any stage (cases 3 and 4), then it indicates the need for further decomposition and hence, we seek a new threshold for the sub-image $F_{01}(x,y)$. We accept the partition, F_{01} when both local and global fits satisfy the criterion C.

A new threshold s_1 divides the image F_{01} into $F_{011}(x,y)$ and $F_{012}(x,y)$. The gray levels in $F_{011}(x,y)$ extend from zero to s_1 while in $F_{012}(x,y)$ they extend from s_1+1 to s. In other words, the gray level bands are $[0, s_1]$ and $[s_1, s]$ respectively for $F_{011}(x,y)$ and $F_{012}(x,y)$. The image $F_{02}(x,y)$ may likewise be examined and segmented if needed. The segmentation, therefore, follows a binary tree structure as shown in Fig. 2.

The criterion C plays a crucial role in the determination of polynomial orders. If the segmented regions are more or less uniform, then low-order polynomials will fit the data reasonably well. However, if the approximation criterion C is very strict and if the spatial distribution of gray values over a region deviates from uniformity, higher-order polynomial will be required to justify the fit. This will result in better reconstruction of image at the cost of compression ratio. Hence, the choice of C should be made based on a compromise between the quality of reconstructed image and the compression ratio. Sections 2.2–2.4 provide details of approximation along with a new approach for the determination of polynomial order. In most of the cases, order is seen to be 2 but it can go upto 3 or 4 depending on variations in the segmented regions and the criterion C.

2.2. Approximation problem

First of all, we formulate the approximation problem using Bezier–Bernstein polynomial and then we address the issue of polynomial order determination. We have chosen the Bezier–Bernstein polynomial because our segmentation algorithm is basically designed for image compression. Bezier–Bernstein polynomial provides a number of merits during reconstruction. However, one can also use other functions.

The Bezier–Bernstein surface is a tensor product surface and is given by

$$\begin{aligned} s_{pq}(u,v) &= \sum_{r=0}^{p}\sum_{z=0}^{q}\phi_{rp}(u)\phi'_{zq}(v)V_{rz} \\ &= \sum_{r=0}^{p}\sum_{z=0}^{q}B_{rp}D_{zq}u^r(1-u)^{p-r}v^z(1-v)^{q-z}V_{rz}, \end{aligned} \tag{13}$$

where $u, v \in [0,1]$ and $B_{rp} = p!/((p-r)!r!)$, $D_{zq} = q!/((q-z)z!)$. p and q define the order of the Bezier–Bernstein surface.

To approximate an arbitrary image surface $f(x,y)$ of size $M\times M$, $f(x,y)$ should be defined in

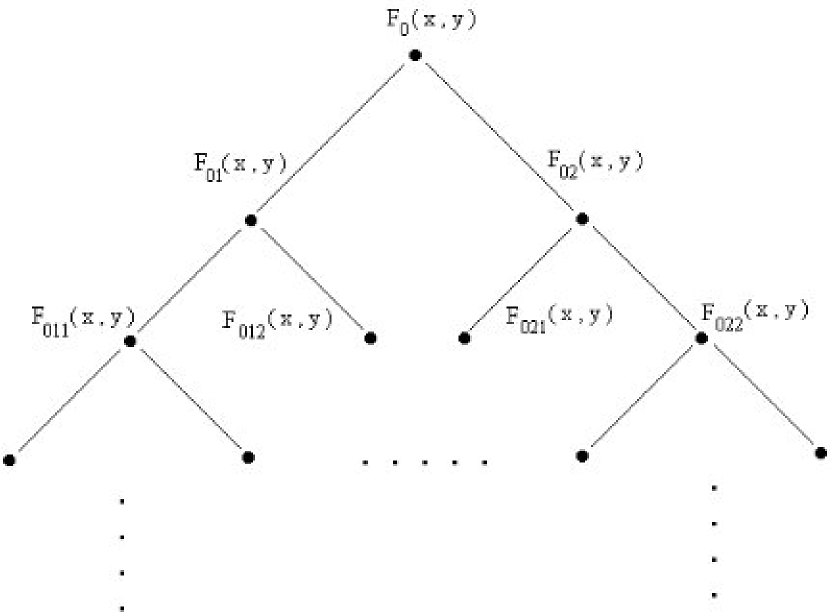

Fig. 2. Binary tree structure for hierarchical segmentation.

terms of a parametric surface (here s_{pq}) with the parameters u, v in $[0, 1]$. Therefore, the function $f(x, y)$ can be thought in terms of $g(u, v)$, where $u = (i-1)/(M-1)$; $i = 1, 2 \ldots, M$ and $v = (j-1)/(M-1)$, $j = 1, 2 \ldots, M$.

We choose the weighted least square technique for estimation of parameters V_{rz} to be used for reconstruction of the decoded surface. Although, the total square error for the conventional unweighted least square approximation may be less than that for the weighted least square, the approximation produced by the latter may be psychovisually more appealing than that by the former, provided weights are chosen judiciously. For an image, edge points are more informative than the homogeneous regions. Edges are the distinct features of an image. Thus, edges should be given more emphasis while approximating an image patch and this can be done through weighted least square.

Thus, the weighted squared error can be written as

$$E^2 = \sum_u \sum_v \left[W(u,v) \left(g(u,v) - s_{pq}(u,v) \right) \right]^2 = \sum_u \sum_v \left[W(u,v) \left(g(u,v) - \sum_{r=0}^{p} \sum_{z=0}^{q} \phi_{rp}(u) \phi'_{zq}(v) V_{rz} \right) \right]^2, \tag{14}$$

where $W(u, v)$ is the weight associated with the pixel corresponding to (u, v). For $p = q$, the surface $s_{pq}(u, v)$ is defined on a square support. Since $W(u, v)$ is the weight associated with each pixel, it can be considered constant for that pixel. Therefore, one needs to find out the weight matrix before solving equations for the weighted least square. Once $W(u, v)$ is known, these equations reduce to a system of linear equations and can be solved by any conventional technique.

We emphasize that for order determination we use the unweighted approximation scheme.

2.3. Polynomial order determination

The order of the polynomial can be determined using either the classical approach or the image quality index (IQI) (Biswas et al., 1994). Since IQI reflects the average contrast (with respect to background) per pixel in the image, if the original and approximated image have nearly the same IQI then the approximated image is expected to preserves the boundary contrast in the average sense. We, therefore, use very small ΔIQI between the input and approximated sub-images as an indicator of the adequacy of the polynomial order. In order to determine optimal polynomial, we increase the order of the polynomial unless the following condition is satisfied

$$|(\mathrm{IQI})_{\mathrm{input}} - (\mathrm{IQI})_{\mathrm{approximated}}| \leqslant \epsilon_a, \tag{15}$$

where ϵ_a is a small positive number.

To calculate IQI we find, first of all, the total contrast K of the image. For an $M \times N$ image, K may be defined as

$$K = \sum_{i=1}^{M} \sum_{j=1}^{N} c_{ij}. \tag{16}$$

The contrast c_{ij}, at the pixel position (i, j) can be written using the concept of psycho-visual perception as (Hall, 1979)

$$c_{ij} = \frac{|B - B_{ij}|}{B} = \frac{|\Delta B|}{B}, \tag{17}$$

where B is the immediate surrounding luminance of the (i, j)th pixel with intensity B_{ij}. Eqs. (16) and (17) reveal that the contrast of pixels in a perfectly homogeneous region is zero everywhere except near the boundary points. The contribution to K of the image, therefore, comes mainly from its noisy pixels and contrast regions (edge points). Thus the image quality index or the average contrast per pixel is defined as

$$\mathrm{IQI} = \frac{K}{n_k}, \tag{18}$$

where $n_k = MN - n_h$, n_k = total number of significant contrast points, n_h = total number of significant homogeneous points and MN = number of pixels in the image. Note that the average is taken over only those pixels which mainly contribute to the contrast measure, K; the pixels of homogeneous regions being least contributory have been discarded.

To find out n_h we define the homogeneity, h_{ij} of the (i,j)th pixel as

$$h_{ij} = \frac{\sum_{r=1}^{8} \exp -|B_{ij} - B_r|}{8}, \tag{19}$$

where B_r indicates the intensity of a background pixel in the 3×3 neighborhood, $N_3(i,j)$, of (i,j). From Eq. (19), it is seen that when each background pixel is equal to the central pixel, the tiny region around the central pixel is perfectly homogeneous, and the homogeneity measure at the central pixel is equal to unity. For other cases, homogeneity value of a pixel drops down exponentially with its difference from the background intensity.

Therefore, if we compute total homogeneity of an image as

$$H = \sum_{i=1}^{M} \sum_{j=1}^{N} h_{ij}, \tag{20}$$

then the major contribution to H comes only from the pixels which lie in perfectly homogeneous regions. Thus, H will be a good *approximation* to n_h. Therefore,

$$\mathrm{IQI} = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} |\Delta B_{ij}| \Big/ B}{MN - \sum\sum h_{ij}}. \tag{21}$$

Thus, the condition in Eq. (15) follows a psycho-visual criterion. A low value of ϵ_a produces psycho-visually a good quality of image. Note that, for an ordinary least square approximation using polynomial surface, the error over the boundary points normally is higher than that over the interior points. Therefore, any polynomial with order determined relative to an error function measured over the boundary points, is expected to provide approximation good for the interior points.

2.4. Algorithms

Method 1: *Variable order global approximation*. Here we determine the order of the global approximation over data points in each sub-image obtained under different thresholds. A schematic description of the global approximation scheme is given below. We assume that there are k number of thresholds for an image and $N_1, N_2, \ldots, N_k$ are the number of regions in these k sub-images.

Algorithm. *global_approx*$(input\_image, th, \epsilon_a, p)$
begin
step 1. Compute the weights as the gradients of the image;
step 2. Find an acceptable sub-image corresponding to a threshold *th* obtained during segmentation by **Algorithm** *Cond_threshold* (assuming $W(i,j) = 1\ \forall i,j$);
step 3. Find the value of IQI of the sub-image using Eq. (21);
step 4. Set the order of the polynomial, $p = 1$;
it step 5. Approximate the sub-image with weights as computed in step 1;
step 6. Find IQI of the approximated image;
step 7. If $|(\mathrm{IQI})_{\text{sub-image}} - (\mathrm{IQI})_{\text{approximated}}| \leqslant \epsilon_a$ then return p and goto step 8 else set $p = p + 1$ and go to step 5;
step 8. stop;
end.

Method 2: *Variable/fixed-order local approximation*. If the variable order global approximation over sub-images does not provide good approximation for some regions in a sub-image then we do local correction. The global approximation is performed over each of the k sub-images using a variable order polynomial function. The residual error surface patches are computed using the globally approximated surface $s_{pp}(u,v)$ and the original input surface (here, the input sub-image). Let us denote lth error surface patch of the ith sub-image by $e_l^i(u,v)$. Considering N_i error surface patches that need local correction in the ith sub-image, we see that

$$e_l^i(u,v) = g(u,v) - s_{pp}(u,v),$$
$$i = 1,2,\ldots,k \quad \text{and} \quad l = 1,2,\ldots,N_i.$$

Each of these error surface patches is approximated locally using a fixed or variable order polynomial. A schematic description of variable order local surface approximation is given below.

Algorithm. *local_approx*$(input\_image, th, \epsilon_a, q, p)$
begin
step 1. find the most dispersed region, Ω_k in the input_image; find the residual error surface for it with respect to order q;
step 2. find p using the **Algorithm** *global_approx* $(\Omega_k, th, \epsilon_a, p)$;
step 3. if $p \geqslant q$, a preassigned positive integer then go to step 4 else assign an index for the region and return p;
step 4. stop;
end;

To summarize, our scheme is a two-stage process. In stage 1, we first determine a threshold. This threshold partitions an image into two sub-images, F_{01} and F_{02}. We determine the order of a polynomial minimizing *unweighted* least square error for approximating a sub-image F_{01}. If the order of the polynomial is less than a predefined order, say, q then we accept the partition F_{01} else we do a local correction for one or more regions. Local correction is always with respect to the global surface of order q. If the global approximation together with local correction(s) is all right, then we accept the sub-image, F_{01} else we compute a new threshold to subdivide F_{01} into F_{011} and F_{012}. The process goes on subdividing the sub-images hierarchically until all of them are approximated by *global_approx* and *local_approx*. The same is also true for F_{02}. The segmentation algorithm may produce some small isolated patches. After the partition of the entire image, all single pixel and small regions or patches are merged to the neighboring regions depending on some criteria which are described in Section 2.5. Note that all approximations in stage 1 are unweighted, i.e., $W(i,j) = 1\ \forall i,j$ in approximation algorithms. In stage 2, for encoding one can approximate the sub-images minimizing a *weighted* least square error with a polynomial of the same order as determined in stage 1. The same order can be used because the order (global and also local) of a sub-image or the nature of approximation is not expected to change due to merging of small regions. However, one can once again find the order of approximation before encoding.

2.5. Merging of small regions

For better segmentation and compression, we like to merge small non-informative regions. This raises two issues: which regions are to be merged and where are to be merged. In order to detect regions of small size for possible merge to one of its neighboring regions, we define a merge index (MI) as the ratio of a measure of within region interactions to that of between-regions interactions. We assume that for a non-trivial region, the within-region interaction should be more than that across the boundary, i.e., $\mathrm{MI} > 1$. A very simple measure of within-region interaction is the number of transitions within the region. Similarly, the between-region interaction can be defined as the number of transitions across the border of the region. Thus, MI can be computed as

$$\mathrm{MI} = (\text{Number of transitions within a region}) / (\text{Number of transitions across the border of the region}).$$

Note that, MI cannot be computed directly from the co-occurrence matrix discussed earlier because more than one isolated regions may contribute to the computation of t_{ij} for a particular (i,j); and in the present context we need to consider only the transitions with respect to one region. This is a very simple, yet effective measure of interaction. Other measures can also be used.

Small regions, detected by MI are the potential candidates for merge and they are merged if the magnitude of the average gradient computed over their region boundaries is less than a preassigned positive value. This criterion will avoid merging small but informative regions. High contrast small regions are usually informative, e.g., the white spot in the eye ball in a face image. The average gradient over a region, say Ω_1 may be computed as

$$\bar{G} = \sum_{(i,j) \in \partial\Omega_1} \frac{G(i,j)}{p}, \tag{22}$$

where p is the perimeter of the region Ω_1 and $G(i,j)$ is the gradient at the position (i,j). The average gradient over other regions can likewise be computed. We have used the following gradient functions. Let $g_{i,j}$ and $g_{k,l}$ be two adjacent pixels belonging to two different regions, say, Ω_i and Ω_k then

$$G(i,j) = \max |g_{i,j} - g_{k,l}| \quad k \in \Omega_k,\ k \in \mathcal{N}_3(i,j), \tag{23}$$

where $N_3(i,j)$ is the 3×3 neighborhood of (i,j). Note that we are not rechecking the segmentation criteria because we are merging small regions with low gradients across the boundary positions. It is expected that the condition will be satisfied and our computational experience indeed supports this fact. However, to ensure the validity of the condition one can check once more the thresholding after merging.

Single pixel merge: Sometimes, single pixel region can occur in a thresholded image. This is merged to the neighboring region having the closest gray value in the 3×3 neighborhood of the single pixel region.

3. Evaluation of segmentation

For the evaluation of segmentation, we focus our attention to region homogeneity and contrast along the boundary points. A good segmentation technique should create homogeneous regions or patches with high contrast at the inter-region boundaries. Note that the type of mergings we have adopted should have very little effect on the overall contrast of the image. We use the following objective measures for quantitative evaluation of segmentation.

3.1. Correlation

Correlation has already been used as a criterion for gray level thresholding and evaluation (Brink, 1989). Here, we use it to examine the gray level similarity between the segmented region/patches and the original image. Consider the segmented image where all patches under respective thresholds are replaced by their average value. The correlation between the segmented and input images provides an idea about how a segmented patch is nearer to the corresponding region in the original input image. For a good segmentation, the correlation coefficient between the two images should be very high. However, if the segmented patches are not homogeneous, i.e., if they have edges in them, the variance of the corresponding regions would be high and as a result the correlation coefficient would be low. Thus correlation, between the two different images; input and segmented, can be an useful measure to evaluate the quality of segmentation. The correlation coefficient can be calculated in the following way.

The coefficient of correlation ρ_{xy} for two sets of data $X = \{x_1, x_2, \ldots, x_N\}$ and $Y = \{y_1, y_2, \ldots, y_N\}$ is given by

$$\rho_{xy} = \frac{\frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x}\bar{y}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N} y_i^2 - \bar{y}^2}}, \tag{24}$$

where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ and $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$. The correlation coefficient, ρ_{xy} takes on values from $+1$ to -1, depending on the type and extent of correlation between the sets of data.

3.2. Contrast

Another requirement for a good segmentation is that the contrast at inter-region boundaries must be very high compared to that for the interior points. This criterion immediately suggests that the average contrast, i.e., contrast per pixel, say $\bar{K}_b$, of all inter-region boundary points in all sub-images should be high compared to that (say, $\bar{K}_\Omega$) over all points enclosed within the boundaries. Therefore,

$$\bar{K}_b \gg \bar{K}_\Omega.$$

The contrast c_{ij}, at the pixel position (i,j) can be computed as in Eq. (17) which we repeat here as

$$c_{ij} = \frac{|B - B_{ij}|}{B} = \frac{|\Delta B|}{B}, \tag{25}$$

where B is the immediate surrounding luminance of the (i,j)th pixel with intensity B_{ij}.

Let SB be the set of all boundary points and SI be the set of all interior points ($SB \cup SI = F$, $SB \cap SI = null$ set). Contrast to all boundary points, K_b and that of interior points, K_Ω are, therefore,

$$K_b = \sum_{(i,j)\in SB} c_{ij} \quad \text{and} \quad K_\Omega = \sum_{(i,j)\in SI} c_{ij}.$$

Note that K_Ω is an indicant of homogeneity within-regions – lower the value of K_Ω, higher is the homogeneity. The contrast per pixel, $\bar{K}_b$, of all inter-region boundary points and that over all points enclosed within the boundaries, $\bar{K}_\Omega$ can be obtained dividing K_b by the number of boundary points and K_Ω by the number of interior points.

4. Comparison with multi-level thresholding algorithms

Since the co-occurrence matrix contains information regarding the spatial distribution of gray levels in the image, several workers have used it for segmentation. For thresholding at gray level 's', Weszka and Rosenfeld (1978) defined the busyness measure as follows:

$$\text{Busy}(s) = \sum_{i=0}^{s} \sum_{j=s+1}^{L-1} t_{ij} + \sum_{i=s+1}^{L-1} \sum_{j=0}^{s} t_{ij}. \tag{26}$$

The co-occurrence matrix used in (26) is symmetric. For an image with only two types of regions, say, object and background, the value of 's' which minimizes Busy(s), 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 (1983) gave a measure which they called "conditional probability of transition" from one region to another as follows. If the threshold is at 's', the conditional probability of transition from the region $[0, s]$ to $[s+1, L-1]$ is

$$P_1 = \frac{\sum_{i=0}^{s} \sum_{j=s+1}^{L-1} t_{ij}}{\sum_{i=1}^{s} \sum_{j=0}^{s} t_{ij} + \sum_{i=0}^{s} \sum_{j=s+1}^{L-1} t_{ij}}, \tag{27}$$

and the conditional probability of transition from the region $[(s+1), (L-1)]$ to $[0, s]$ is

$$P_2 = \frac{\sum_{i=s+1}^{L-1} \sum_{j=0}^{s} t_{ij}}{\sum_{i=s+1}^{L-1} \sum_{j=s+1}^{L-1} t_{ij} + \sum_{i=s+1}^{L-1} \sum_{j=0}^{s} t_{ij}}. \tag{28}$$

$p_c(s)$, the conditional probability of transition across the boundary is then defined as

$$p_c(s) = (P_1 + P_2)/2. \tag{29}$$

Expressions (27)–(29) suggest that 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. (1985) also used the co-occurrence matrix for thresholding. They defined an average contrast measure as

$$\text{AVC}(s) = \frac{\sum_{i=0}^{s} \sum_{j=s+1}^{L-1} t_{ij} * (i-j)^2}{\sum_{i=0}^{s} \sum_{j=s+1}^{L-1} t_{ij}} + \frac{\sum_{i=s+1}^{L-1} \sum_{j=0}^{s} t_{ij} * (i-j)^2}{\sum_{i=s+1}^{L-1} \sum_{j=0}^{s} t_{ij}}. \tag{30}$$

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

In the computation of t_{ij}, they considered only vertical transitions in the downward direction.

4.1. Results

In order to evaluate the quality of segmentation produced by the proposed algorithm and to compare it with those of the above three different algorithms, we examine in Table 1 the objective measures discussed in the previous section. We have used two 32-level images (Figs. 3 and 4), each of size 64 × 64. Fig. 3(a) is the Lincoln image while Fig. 4(a) is the Biplane image. Table 1 shows the values of different objective measures in conjunction with the total number of regions or patches, say N_Ω, produced by different segmentation techniques for the images. Note that the number of regions is an important parameter to justify goodness of segmentation. For the Lincoln image, the number of segmented regions obtained by the proposed algorithm is almost one-fourth of those obtained by the other algorithms and for Biplane image the number of regions is roughly half of those produced by the algorithms of Rosenfeld and Kak (1982), Deravi and Pal (1983) and Chanda et al. (1985), respectively.

Table 1
Evaluation of different segmentation algorithms

Objective measure	Proposed	Weszka and Rosenfeld (1978)	Deravi and Pal (1983)	Chanda et al. (1985)
Lincoln image				
Number of regions, N_Ω	52	187	192	189
Correlation	0.978784	0.987864	0.987307	0.990799
Boundary contrast/pixel, $\bar{K}_b$	0.204	0.202	0.200	0.194
Region contrast/pixel, $\bar{K}_\Omega$	0.0294	0.0257	0.0258	0.0293
Biplane image				
Number of regions, N_Ω	35	59	59	76
Correlation	0.988588	0.989192	0.989192	0.988381
Boundary contrast/pixel, $\bar{K}_b$	0.1499	0.1866	0.1866	0.1782
Region contrast/pixel, $\bar{K}_\Omega$	0.0151	0.0144	0.0144	0.0150

Usually, with the increase in number of regions, correlation is expected to increase. The segmentation of both Lincoln and Biplane images supports this fact. But even with a much smaller number of regions for both the images, produced by the proposed scheme, the correlation values are comparable to those for the segmented images obtained from other algorithms. This indicates successful merging of small regions to the proper neighboring regions. Also, due to the merging the homogeneity of the segmented regions is expected to increase. For good segmentation, this homogeneity should be very high. In other words, the average contrast $\bar{K}_\Omega$ within a region should be low. The parameter region contrast/pixel, $\bar{K}_\Omega$ shows that the average homogeneity is reasonably good. Finally, the average boundary contrast $\bar{K}_b$, for both images is very much comparable to all the cases. Thus, the proposed scheme for segmentation is better than other methods and provides advantage from the standpoint of compression. Different segmented images along with the input are shown in Fig. 3((a)–(e)) and Fig. 4((a)–(e)). For a better display of segmented regions, all segmented images are stretched over a gray scale of [0–255].

5. Conclusions and discussion

We have developed a new segmentation scheme, keeping in mind its use in image data compression. The segmentation strategy uses the multi-level thresholding based on conditional entropy. It partitions an image hierarchically and merges a small region efficiently. Evaluation of the proposed segmentation scheme has been performed and its result has been compared with those of several existing multi-thresholding schemes.

The proposed algorithm in this paper shows the possibility of globally approximating many segmented regions or patches by a single polynomial function. In other words, attempts have been made to model different regions in an image by a single polynomial surface. For this, all such regions should have similar gray levels. The segmented regions to be approximated by a single polynomial can be extracted under a single threshold. Thresholding-based segmentation thus provides an advantage over split and merge technique of segmentation (Pavlidis, 1982). The latter does not provide any group of patches or regions of similar gray levels located at different places in an image at a time. It is, therefore, preferable to choose a thresholding technique of segmentation for coding application. Because under such segmentation, a set of approximation parameters can represent many regions. This set of parameters represents a single surface on which different regions are situated at different locations. Hence we do not need to code all regions separately for their gray information. This is an important reason responsible for providing advantage to image compression. However, the gray level distribution over some of the image surface patches may be such that the global approximation is not adequate for them. We call such patches, under a given threshold, *busy* patches. To overcome this difficulty, a lower order (compared to that of the global

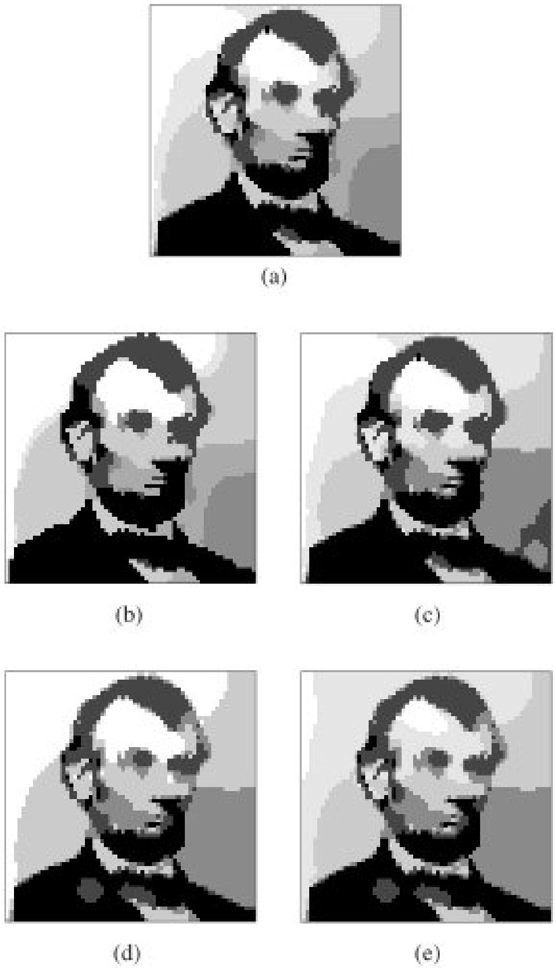

Fig. 3. (a) Input Lincoln image; (b) segmented image by the proposed method; (c) segmented image by Chanda et al. (1985); (d) segmented image by Weszka and Rosenfeld (1978); (e) segmented image by Deravi and Pal (1983).

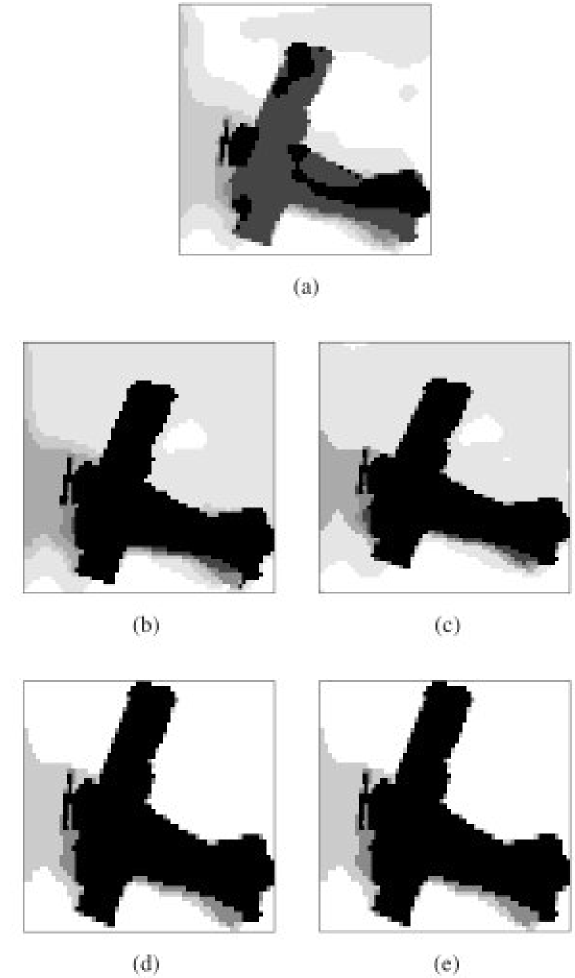

Fig. 4. (a) Input image of a biplane; (b) segmented image by the proposed method; (c) segmented image by Chanda et al. (1985); (d) segmented image by Weszka and Rosenfeld (1978); (e) segmented image by Deravi and Pal (1983).

approximation) polynomial function can be used for local approximation of each of the residual surface patches in the sub-image. Therefore, a sub-image can be reconstructed using the global surface along with the local residual surfaces for the busy patches if they are really present. Such a hybrid approximation scheme helps to improve the compression ratio. Note that exactly the same kind of approximation is used to guide the segmentation process which ensures that the extracted sub-images can be modeled by low-order polynomials resulting in better compression.

To visualize more clearly the advantage provided by the proposed algorithm to image compression we consider the following example.

Suppose in a threshold band limited sub-image $F(x, y)$ we have N surface patches, then for the local quadratic approximation one requires $6N$ coefficients. On the other hand, if we have the global quadratic approximation of the sub-image and local planar approximation of the residual surface patches, the total number of coefficients is $3N + 6$. For an improvement in compression ratio of the global–local approximation over the conventional local approximation we must have $6N > 3N + 6$, i.e., $N > 2$. This implies a positive gain in storage if the sub-image has more than two surface patches which usually is always the case. Thus, it is evident that for polynomial approximation, we need less number of bits for any segmentation-based lossy image Compression technique where regions or patches are approximated separately. Compression factor, as a result, would improve (assuming the same contour coding scheme as in the concerned method).

Performance of a compression algorithm based on the proposed segmentation scheme will be reported in a forthcoming paper.

References

Abutaleb, A.S., 1989. Automatic thresholding of graylevel pictures using two-dimensional entropy. Computer Graphics, Vision and Image Processing 47, 22–32.

Biswas, S., Pal, N.R., Pal, S.K., 1994. A quantitative index for termination of iterative image smoothing algorithms. In: Proc. 3rd Internat. Conf. on Automation, Robotics and Computer Vision, pp. 1107–1111.

Brink, A.D., 1989. Grey-level thresholding of images using a correlation criterion. Pattern Recognition Letters 9, 335–341.

Carlsson, S., 1988. Sketch based coding of gray level images. Signal Processing 15 (1), 57–83.

Chanda, B., Choudhuri, B., Majumder, D.D., 1985. Minimum error thresholding. Pattern Recognition Letters 3 (4), 243–251.

Deravi, F., Pal, S.K., 1983. Graylevel thresholding using second-order statistics. Pattern Recognition Letters 1 (5), 417–422.

Fu, K.S., Mui, J.K., 1981. A survey of image segmentation. Pattern Recognition 13.

Gonzalez, R.C., Wintz, P., 1977. Digital Image Processing. Addison-Wesley, Reading, MA.

Hall, E.H., 1979. Computer Image Processing and Recognition. Academic Press, New York.

Haralick, R.M., Shapiro, L.G., 1985. Image sementation techniques. Computer Vision, Graphics and Image Processing 29, 100–132.

Kapur, J.N., Shaoo, P.K., Wong, A.K.C., 1985. Gray level picture thresholding using the entropy of histogram. Computer Vision, Graphics and Image processing 29, 273–285.

Kitler, J., Illingworth, J., 1986. Minimum error thresholding. Pattern Recognition Letters 19 (1), 97–108.

Kunt, M., Benard, M., Leonardi, R., 1987. Recent results in high compression image coding. IEEE Transactions on Circuits and Systems 34 (11), 1306–1336.

Kunt, M., Ikonomopoulos, A., Kocher, M., 1985. Second-generation image coding techniques. In: Proc. IEEE 73, 549–574.

Pal, N.R., Bhandari, D., 1993. Image thresholding. Signal Processing 33 (2), 139–158.

Pal, N.R., Pal, S.K., 1989a. Entropic thresholding. Signal Processing 16 (2), 97–108.

Pal, N.R., Pal, S.K., 1989b. Object background segmentation using new definition of entropy. In: Proc. IEEE, Vol. part E, pp. 284–295.

Pavlidis, T., 1977. Structural Pattern Recognition. Springer, New York.

Pavlidis, T., 1982. Algorithms for Graphics and Image Processing. Springer, New York.

Pun, T., 1980. A new method for gray level picture thresholding using the entropy of the histogram. Signal Processing 2 (3), 223–237.

Pun, T., 1981. Entropic thresholding a new method of image segmentation. Computer Graphics and Image Processing 16, 210–239.

Rosenfeld, A., Kak, A.C., 1982. Digital Picture Processing. Academic Press, Orlando, FL.

Shannon, C.E., Weaver, W., 1949. The Mathematical Theory of Communication. University of Illinois Press, Urbana.

Shen, L., Rangayyan, R.M., 1997. A segmentation-based lossless image coding method for high resolution medical image compression. IEEE Transactions on Pattern Analysis and Machine Intelligence 16, 301–307.

Weszka, J.S., 1978. A survey of threshold selection techniques. Computer Graphics and Image Processing 7, 259–265.

Weszka, J.S., Rosenfeld, A., 1978. Threshold evaluation techniques. IEEE Transactions on Systems, Man and Cybernetics 8, 622–629.