INFORMATION SCIENCES 60, 107–135 (1992) 107

Histogram Thresholding by Minimizing Graylevel Fuzziness*

CHIVUKULA A. MURTHY

and

SANKAR K. PAL

Electronics and Communication Sciences Unit, Indian Statistical Institute, 203 B.T. Road, Calcutta 700035, India

ABSTRACT

The problem of Histogram sharpening and thresholding by minimizing greyness ambiguity using the measures of fuzziness in a set is considered. This work provides a complete mathematical formulation of the said problem in order that both fuzzy and crisp segmentation result. The criteria regarding the choices of membership functions and the window sizes are established. They involve symmetry in ambiguity around the crossover point and bound functions for restricting the variation of membership function values. The segmentation thus obtained provides not only flexibility in methodology but also effective results. Experimental results further establish the same conclusions.

1. INTRODUCTION

A measure of ambiguity (fuzziness) in grey level of an image is seen to be provided [1] by the terms index of fuzziness [2], entropy [3], and index of nonfuzziness [4]. Index of fuzziness reflects the average amount of ambiguity present in an image X by measuring the distance ("linear" and "quadratic" corresponding to linear index of fuzziness and quadratic index of fuzziness) between its fuzzy property μ_X and the nearest two-level property $\mu_{\underline{X}}$. In other words, the distance between grey tone image and its nearest two-tone version. The term "entropy," on the other hand, uses Shannon's function but its meaning is quite different from classical entropy because no probabilistic concept is needed to define it. The index of nonfuzziness, as its name implies,

*Communicated by Abraham Kandel.

measures the amount of nonfuzziness (crispness) in μ_x by computing its distance from its complement version.

Since these terms basically reflect the measure of closeness of a greytone image to its two-tone version, they provide a quantitative measure of image ambiguity [1] when the crossover point is set to a predetermined value. Modification of crossover point will result in variation in these values and so a set of minima may be obtained corresponding to the optimum threshold levels of the image.

The present work provides a mathematical formulation to illustrate an application of the aforesaid measures to automatic grey level thresholding (image segmentation). Given an image, the algorithm ensures, without referring to its histogram, to detect the valleys corresponding to the optimum thresholds by minimizing the greyness ambiguity as reflected by the said measures.

The mathematical framework considers a nondecreasing fuzzy membership function g taking values $0 \leq g \leq 1$ over an interval of length c (window size). By changing crossover point (i.e., by moving g) over the entire dynamic range of the image, keeping c fixed, the values of the fuzzy measures are computed. It is then established that, by suitably taking the value of c, this operation removes the local variation and hence sharpens the histogram, thus enabling one to detect the valley point by minimizing fuzziness in the vicinity of valley of the input histogram. The image μ_X generated by that membership function g corresponding to the valley point represents an optimum fuzzy segmentation of X. Here the membership function of a pixel denotes the degree of its belonging to the object (or the subset "bright image"). This is optimum in the sense that for any other choice of crossover point of g the ambiguity of μ_X will be greater than this. To obtain a nonfuzzy segmented version, one may consider the corresponding crossover point as the threshold.

It is to be mentioned here that a similar concept was used earlier experimentally by Pal et al. [5] and Pal and Rosenfeld [6] considering only Zadeh's standard S function [7] as g in extracting a fuzzy subset "bright image" from the image. But the authors did not provide any mathematical basis of their findings. For example, the limitation of using any other type of g function is neither theoretically nor even experimentally justified. The observations on the choice of c (which is critical for detecting valleys) was provided only experimentally.

The work outlined in this paper mathematically establishes the said experimental results and provides answers to the above queries. The framework takes into consideration all possible membership functions and histograms. The relation between c and the length of the interval between two peaks of a histogram is established. The effect of variation of membership function (i.e., the limitation on the choice of g) on the results is mathematically described. It is then found out that the requisite membership function g may be confined within the bounds of Murthy and Pal [8] and it possesses symmetry in ambiguity around the crossover point. With this, the method is therefore seen to be flexible enough in selecting its input membership function, keeping the output satisfactory. The results are further demonstrated on a few input images.

Besides giving a generalization of the earlier works for image sharpening/thresholding my minimizing fuzziness in an image, this work can be regarded as a completion of the same.

2. GREYNESS AMBIGUITY AND THRESHOLD SELECTION

Let X be an image of $L + l$ levels, M rows, and N columns and let μ be a membership function defined on levels. Let $\mu(x_{mn})$ denote the grade of possessing some property (e.g., brightness) by the (m, n) pixel of intensity x_{mn}, $m = 1, \ldots, M$ and $n = 1, \ldots, N$. The index of fuzziness $[\gamma(X)]$, entropy $[E(X)]$, and index of nonfuzziness $[\eta(X)]$ are defined below [1]:

$$\gamma(X) = \frac{2}{MN} \sum_m \sum_n \operatorname{Min}[\mu(x_{mn}), (1 - \mu(x_{mn}))]$$

$$= \frac{2}{MN} \sum_m \sum_n |\mu(x_{mn}) - \mu_X(x_{mn})|, \tag{1a}$$

where μ_X represents the nearest two-tone version of μ,

$$E(X) = \frac{1}{MN \ln 2} \sum_m \sum_n Sn(\mu(x_{mn})) \tag{1b}$$

with Shannon's function

$$Sn(\mu(x_{mn})) = -\mu(x_{mn}) \log \mu(x_{mn})$$

$$-(1 - \mu(x_{mn})) \log(1 - \mu(x_{mn})), \tag{1c}$$

$$\eta(X) = \frac{1}{MN} \sum_m \sum_n [1 - |2\mu(x_{mn}) - 1|]. \tag{1d}$$

Intuitively ambiguity in greyness should be maximum when $\mu(x_{mn}) = 0.5$ and it should decrease as $\mu(x_{mn})$ moves away from 0.5. The above-mentioned measures possess this property and hence they can be considered to represent the greyness ambiguity in X.

Observe that $\eta(X)$ is same as $\gamma(X)$ because of the reason mentioned below:

(a) Let $\mu(x_{mn}) \leqslant 1/2$. Then $1 - |2\mu(x_{mn}) - 1| = 2\mu(x_{mn})$.
(b) Let $\mu(x_{mn}) \geqslant 1/2$. Then $1 - |2(\mu(x_{mn})) - 1| = 2(1 - \mu(x_{mn}))$.

Let $f(l)$ denote the number of occurrences of the level l. Equations (1) can then be written as

$$\gamma(X) = \frac{2}{MN} \sum T(l) f(l) \tag{2a}$$

with

$$T(l) = \text{Min}[\mu(l), 1 - \mu(l)] \tag{2b}$$

and

$$E(X) = \frac{1}{MN \ln 2} \sum Sn(\mu(l)) f(l) \tag{3}$$

The concept of using index of fuzziness for threshold selection is described below. A similar argument holds for entropy also. Let Zadeh's standard S function [7] be considered μ here.

$$\begin{aligned}
\mu(x_{mn}; p, q, r) &= 0 && \text{if } x_{mn} \leqslant p \\
&= 2[(x_{mn} - p)/(r - p)]^2 && \text{if } p \leqslant x_{mn} \leqslant q \\
&= 1 - 2[(x_{mn} - r)/(r - p)]^2 && \text{if } q \leqslant x_{mn} \leqslant r \\
&= 1 && \text{if } x_{mn} \geqslant r
\end{aligned} \tag{4}$$

with $q = (p + r)/2$ and $\Delta q = r - q = q - p$. The parameter q is the crossover point. The window length $= r - p = 2\Delta q$.

Let us, for example, consider the object and background segmentation of a bimodal histogram. Now, for an image X, the fuzzy measures basically compute the distance between its brightness property μ_X and its nearest two-tone property $\mu_{\mathbf{X}}$. Since, X is dependent on the position of the crossover point q, a proper selection of q (and hence the membership function) may therefore be obtained which will result in minimum value of these measures γ and E. This minimum value corresponds to appropriate segmentation of the image and q may be taken as optimum threshold. This is optimum in the sense that, for any other choice of q, the γ or E measure will be greater than this.

The corresponding $\mu(x_{mn})$ plane can be regarded as a fuzzy segmented version of X. For obtaining its nonfuzzy (crisp) version, the crossover point q (having maximum ambiguity) was considered above as the threshold between object and background. The above concept can similarly be extended to a multimodal image where there would be several minima corresponding to different valley points of the histogram.

Algorithm for Greylevel Thresholding:

Fuzzy membership function $\mu(x_{mn}; p, q, r)$ [Equation (4)] is considered. The function μ is shifted over the interval $[0, L]$ by varying p, q, and r but keeping Δq fixed. When Δq is fixed, the whole function μ can be determined uniquely given q. Indices of fuzziness are calculated for every μ, i.e., for every q. The valley points of $\gamma(q)$ are taken to be the detected thresholds (unambiguous valley points) of the histogram of the input image. The algorithm is thus seen to be able to sharpen an input histogram by removing the local variations and ambiguities in the vicinity of its valleys.

In this algorithm $c = 2\Delta q$ is the length of the interval which is shifted over the entire dynamic range. As c decreases, $\mu(x_{mn})$ plane would have more intensified contrast around the crossover point, resulting in a decrease of ambiguity in X. As a result, the possibility of detecting some undesirable thresholds (spurious minima) increases because of the smaller value of Δq.

On the other hand, increase of c results in a higher value of fuzziness and thus leads towards the possibility of losing some of the weak minima.

Though in some earlier works [5, 6] the said concept was used, the mathematical formulation of the problem was not provided. For example, it was reported that if c is greater than the distance between the modes, then the corresponding valley point may be lost [5]. But the mathematical justification of this finding was not given. This is similar to the case with the selection of the membership function where only the function shown in Equation (4) was considered. The consequences of using other types of membership functions are neither mathematically nor experimentally studied.

In Section 3, mathematical formulation of the problem and some of its consequences are stated. The relation between c [window size] and the distance between modes is also established in Section 3. In Section 4, different types of membership functions are taken and the corresponding changes on the thresholds of histogram are discussed. Section 5 provides the experimental results. Discussions and conclusion are given in Section 6.

3. MATHEMATICAL FORMULATION OF HISTOGRAM THRESHOLDING

We shall assume continuous functions for the formulation and proofs. Similar results hold in discrete cases also. The histogram will be represented by f, the membership function by g, and the index of fuzziness by $M_0 H$, where M_0 is a constant [$M_0 = 2/MN$ of Equation (2a)] and H represents the other part of Equation (2a). [The summation sign should be changed by integral because of continuity].

THEOREM 1. *Let* $f = [0, L) \to [0, \infty)$ *be such that (i)* f *is continuous, (ii)* $f(a) = f(b)$, $a < b$ *and* f *has local maximums at* a *and* b, *(iii)* $\mathscr{S}_0 = (a+b)/2$, f *has a local minimum at* $\mathscr{S}_0$, *(iv)* f *is symmetric around* $\mathscr{S}_0$ *in the interval* $[a,b]$ *and (v)* f *is convex in* $[a,b]$. *Let* $g\colon [0,c] \to [0,1]$ *be such that (i)* g *is continuous,* $g(0) = 0$, $g(c) = 1$, *(ii)* g *is monotonically nondecreasing, and (iii)* $g(x) = 1 - g(c-x)$ $\forall x \in [0,c]$. $c > 0$ *is the length of the window. Let* $c < b - a$. *Let* $\partial = (b - a - c)/2$. *Let*

$$H_g(y) = \int_0^{c/2} g(x) f(y - c/2 + x)\, dx + \int_{c/2}^{c} (1 - g(x)) f(y - c/2 + x)\, dx.$$

[Observe that $M_0 H_g(y)$ *gives the index of fuzziness in the interval* $(y - c/2, y + c/2)$*]. Then* $H_g(y) \geqslant H_g(y_0)$ $\forall y \in (y_0 - \partial, y_0 + \partial)$.

Proof. Let $0 < \epsilon < \partial$. Let $y = y_0 - \epsilon$. We shall show that

$$H_g(y) \geqslant H_g(y_0)$$

$$H_g(y) = \int_0^{c/2} g(x) f(y_0 - \epsilon - c/2 + x)\, dx + \int_{c/2}^{c} (1 - g(x)) f(y_0 - c/2 - \epsilon + x)\, dx$$

$$= I_1 + I_2 \text{ (say)}$$

We shall simplify I_2 now with the transformation $x = c - z$:

$$I_2 = -\int_{c/2}^{0} [1 - g(c - z)] f(y_0 - c/2 - \epsilon + c - z)\, dz$$

$$= \int_0^{c/2} g(z) f(y_0 + c/2 - \epsilon - z)\, dz$$

$$= \int_0^{c/2} g(z) f(y_0 - c/2 + \epsilon + z)\, dz$$

($\because f$ is symmetric in $[a,b]$).

So

$$H_g(y) = \int_0^{c/2} g(x) [f(y_0 - \epsilon - c/2 + x) + f(y_0 - c/2 + \epsilon + x)]\, dx$$

$$= 2\int_0^{c/2} g(x) [\tfrac{1}{2} f(y_0 - \epsilon - c/2 + x) + \tfrac{1}{2} f(y_0 + \epsilon - c/2 + x)]\, dx$$

$$\geqslant 2\int_0^{c/2} g(x) f\{\tfrac{1}{2}[y_0 - \epsilon - c/2 + x + y_0 + \epsilon - c/2 + x]\}\, dx$$

$$= 2\int_0^{c/2} g(x) f(y_0 - c/2 + x)\, dx. \tag{5}$$

Now

$$H_g(y_0) = \int_0^{c/2} g(x) f(y_0 - c/2 + x)\, dx + \int_{c/2}^{c} (1 - g(x)) f(y_0 - c/2 + x)\, dx$$

$$= J_1 + J_2.$$

By applying the same calculation of I_2 to J_2, it can be shown that

$$J_2 = \int_0^{c/2} g(x) f(y_0 - c/2 + x)\, dx,$$

i.e., $H_g(y_0) = 2\int_0^{c/2} g(x) f(y_0 - c/2 + x)\, dx \leqslant H_g(y)$ [from (5)].
For $y = y_0 + \epsilon$ when $0 < \epsilon < \partial$, similar proof holds. Hence the theorem. ■

REMARK 1.

(a) Similar proof can be given if the entropy is taken to be the grey level ambiguity measure.

(b) ∂ in the above theorem will give an idea of the length of the interval in which $H_g(y) \geq H_g(y_0)$. For c being close to $b - a$, ∂ will be very small and the valley will be obtained in a smaller interval. In practical cases the valley may become invisible also. The case of $c = b - a$ is tackled below in Note 1.

Note 1.

Let f and g satisfy the same assumptions as in theorems. Let $\epsilon > 0$ be a small quantity $c = b - a$ and $y = y_0 - \epsilon$. So

$$H_g(y) = \int_0^\epsilon g(x) f(y_0 - \epsilon - c/2 + x)\,dx$$

$$+ \int_\epsilon^{c/2} g(x) f(y_0 - \epsilon - c/2 + x)\,dx$$

$$+ \int_{c/2}^{c-\epsilon} (1 - g(x)) f(y_0 - \epsilon - c/2 + x)\,dx$$

$$+ \int_{c-\epsilon}^{c} (1 - g(x)) f(y_0 - \epsilon - c/2 + x)\,dx$$

$$= I_1 + I_2 + I_3 + I_4 \quad \text{(say)},$$

$$H_g(y_0) = \int_0^\epsilon g(x) f(y_0 - c/2 + x)\,dx$$

$$+ \int_\epsilon^{c/2} g(x) f(y_0 - c/2 + x)\,dx$$

$$+ \int_{c/2}^{c-\epsilon} (1 - g(x)) f(y_0 - c/2 + x)\,dx$$

$$+ \int_{c-\epsilon}^{c} (1 - g(x)) f(y_0 - c/2 + x)\,dx$$

$$= J_1 + J_2 + J_3 + J_4 \quad \text{(say)}.$$

Now, as in Theorem 1, I_3 can be proved to be equal to

$$\int_\epsilon^{c/2} g(x) f(y_0 + x + \epsilon - c/2)\,dx.$$

Similar to Theorem 1, $I_2 + I_3$ can be proved to be greater than or equal to

$$2\int_\epsilon^{c/2} g(x) f(y_0 - c/2 + x)\,dx = J_2 + J_3.$$

Now $J_1 + J_4$ can be proved to be equal to $2\int_0^\epsilon g(x) f(y_0 - c/2 + x)\,dx$. I_4 can be shown to be equal to $\int_0^\epsilon g(x) f(y_0 - c/2 + \epsilon + x)\,dx$. Therefore,

$$I_1 + I_4 - J_1 - J_4 = \int_0^\epsilon g(x)\big[f(y_0 - \epsilon - c/2 + x)$$

$$+ f(y_0 - c/2 + x + \epsilon) - 2f(y_0 - c/2 + x)\big]\,dx.$$

If ϵ is sufficiently small, this difference may (not always) become negligible (because g is continuous and f is continuous) which in turn gives $H_g(y_0) \leq H_g(y)$. So for $c = b - a$, practically, it is not always guaranteed that $H_g(y_0) < H_g(y)$. For $c > b - a$, a similar conclusion can be arrived at. So it can be conclusively stated that, for achieving a valley in $H_g(y)$ corresponding to a valley in f, the window length c should be less than the distance between two peaks, if f and g satisfy the conditions of Theorem 1.

In practical problems, the modes of histograms may not be found out exactly. For example, in Figure 2, though it appears that there are two modes, the value of the second mode is not exactly known. In Figure 3, another histogram is shown, where the two modes are known more or less accurately

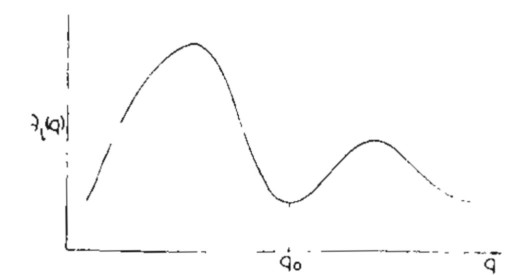


Fig. 1. Graph showing index of fuzziness values vs. crossover points. q_0 is optimal.

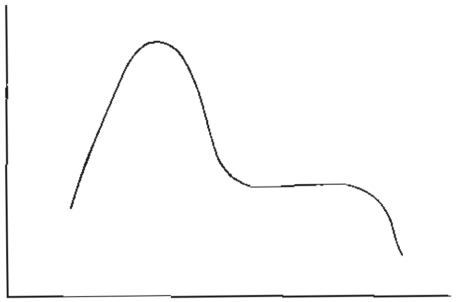


Fig. 2. Histogram in which a mode can't be precisely denoted.

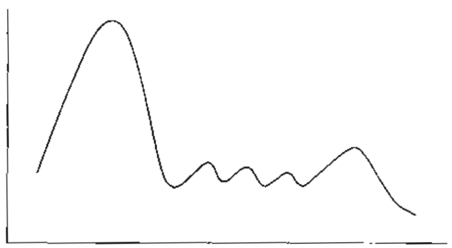


Fig. 3. Histogram where convexity and symmetry properties are not satisfied.

but the convexity and symmetricity properties do not hold and also there are many other local maxima. By using fuzzy membership functions for sharpening the histogram, we would like to remove the redundant local minima of the histogram of Figure 3, so that $H_g(y)$ in Theorem 1 would have one minimum in between the two modes. That is, the membership function should be taken in such a way that $H_g(y)$ should remove "unnecessary" local minima of the histogram. We shall show below that if the value of c is very small, then $H_g(y)$ would give many local minima (Example 1).

EXAMPLE 1. Consider Figure 4 where a histogram f is shown. Though the prominent modes are a_1 and a_2, there are other local maxima, namely a_3, a_4, a_5, and a_7. Let us consider a_1 and a_5. y_0 is the only local minima between a_1 and a_5. f is convex in the interval a_1 and a_5. Let a_6 be such that $f(a_6) = f(a_5)$ and $a_1 < a_6 < y_0 < a_5$. Now suppose that f is symmetric in the interval (a_6, a_5) around y_0. Let $c < a_5 - a_6$. Then any g satisfying the conditions of Theorem 1 would have y_0 as a local minimum. But whether the detection of y_0 as a threshold is desirable or not depends on the problem. For the same histogram shown in Figure 4 and for a two-class problem $H_g(y)$ may produce other local minima if the value of c is not big enough. In practical problems, whether the

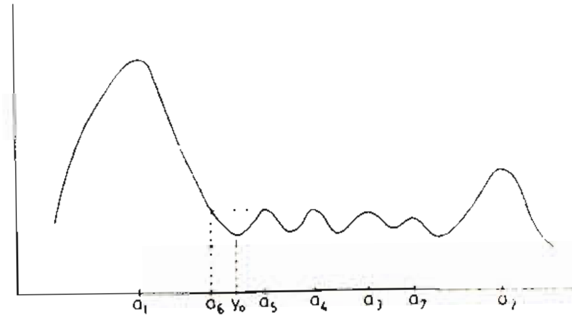


Fig. 4 Histogram of Examples 1 and 2

conditions of convexity and symmetricity are satisfied or not, it is better to take c to be $\geqslant a_5 - a_6$ if the detection of y_0 is to be avoided.

From Example 1, it is apparent that the value of c cannot be very small compared to the difference between the modes. In this section, though all the results are stated for index of fuzziness, the same conclusions would hold for entropy also.

In the next section the effect of various types of membership functions on the valley points are observed.

4. VARIOUS MEMBERSHIP FUNCTIONS AND GREYNESS AMBIGUITY

In the previous section the relation between c and the difference in local maxima is established. In this section, different types of membership functions are examined for the threshold selection using greyness ambiguity. In Example 2, the same histogram of Example 1 is considered to show that some types of membership functions may provide undesirable results.

EXAMPLE 2. The histogram under consideration is the one shown in Figure 4. Consider $g_{\epsilon,\delta}$ as shown in Figure 5 for $\epsilon > 0$ and $\partial > 0$. [Though a specific form of g is presented below, any form with the same idea would suffice for this purpose.]

$$g_{\epsilon,\delta}(x) = \frac{2\epsilon x}{c - a_5 + a_6 + \delta} \qquad \text{for } 0 \leqslant x \leqslant \frac{c - a_5 + a_6 + \partial}{2} = A_1$$

$$= \frac{1}{2} + \frac{(1-2\epsilon)(x - c/2)}{a_5 - a_6 - \partial} \qquad \text{for } A_1 \leqslant x \leqslant \frac{c}{2}$$

$$= 1 - g_{\epsilon,\partial}(c - x) \qquad \text{for } \frac{c}{2} \leqslant x \leqslant c,$$

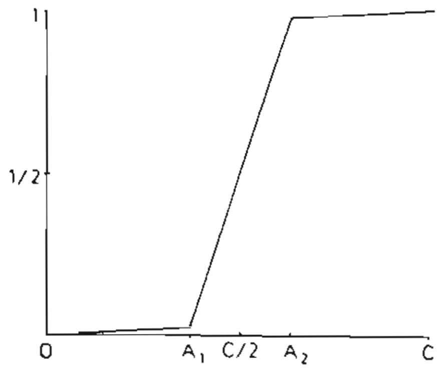


Fig. 5 Membership function of Example 2.

where $0 < \epsilon$ is a small quantity, $0 < \partial < a_5 - a_6$, a_5, and a_6 are as defined in Example 1, c is any positive number, and $A_2 = c - A_1$. The essential differences between g of Example 1 and $g_{\epsilon,\partial}$ of this example are listed below.

	g of Example 1	$g_{\epsilon,\delta}$ of Example 2
1	The length of domain of g is less than $a_5 - a_6$	The length of domain of $g_{\epsilon,\partial}$ is greater than $a_5 - a_6$
2.	Outside the interval of length $a_5 - a_6$, g takes values 0 or 1 so that greyness ambiguity will be zero for those values	Outside the middle interval of length $a_5 - a_6 - \partial$, either Shannon's function or $\text{Min}(g_{\epsilon,\partial}(x), 1 - g_{\epsilon,\partial}(x))$ takes very small values (because ϵ can be made arbitrarily small) so that, after the multiplication with f, the result would be insignificant. That is, $g_{\epsilon,\partial}$ serves the same purpose of g of Example 1. Hence valley y_0 will be detected

Note 2.

In order that y_0 of Figure 4 should not be detected as a valley point of greyness ambiguity function, not only that the value of $c > a_5 - a_6$, but also the membership function $g_{\epsilon,\partial}$ of Example 2 is to be avoided. $g_{\epsilon,\partial}$ has most of its variation concentrated in a small interval in the middle of $[0, c]$ and has little variation in the rest. So this function would not satisfy the bounds of Murthy and Pal [8]. [The bounds of Murthy and Pal are described in the

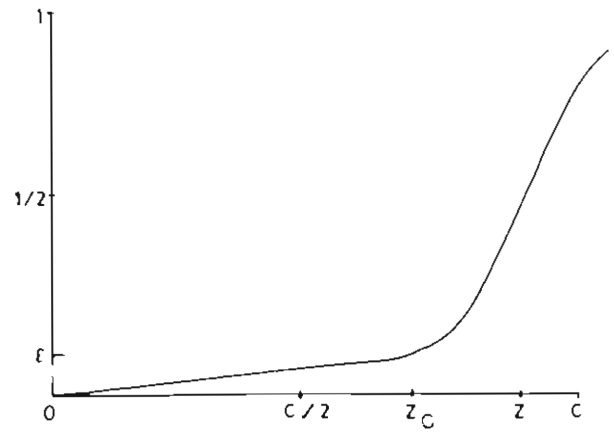


Fig. 6 Membership function of Note 3.

Appendix.] In the practical problems where assumptions of convexity and symmetricity are not satisfied for f, it is imperative that the functions of the sort $g_{\epsilon,\partial}$ are to be avoided.

We will now show that if most of the variation in g is concentrated towards one of the end points of the interval $[0, c]$, it is inadvisable to consider that function [Note 3]. The argument will be similar to that of Example 2.

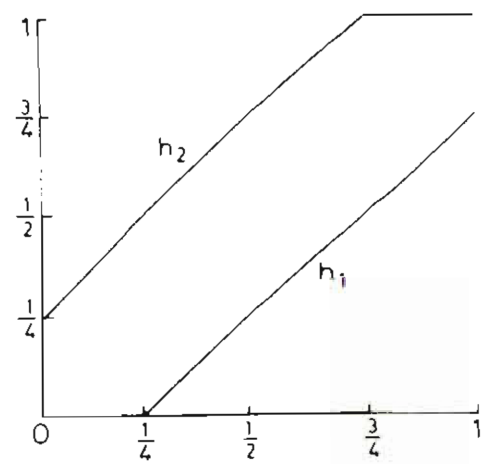
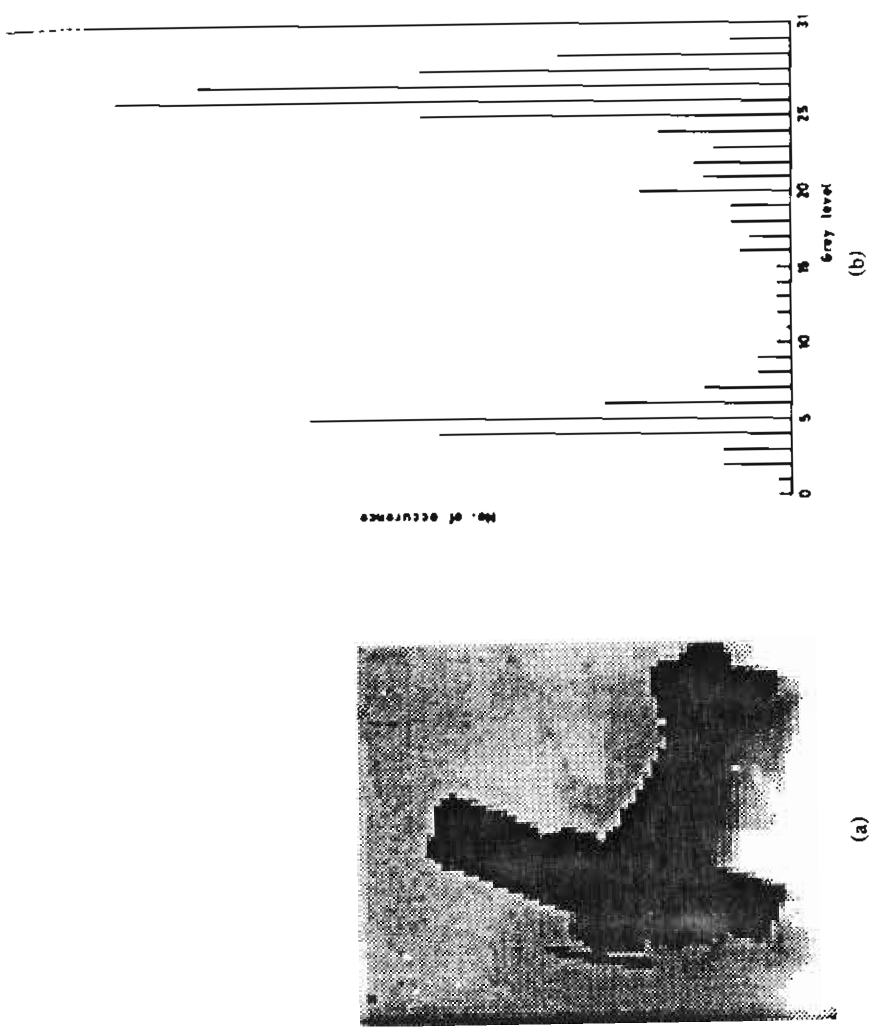
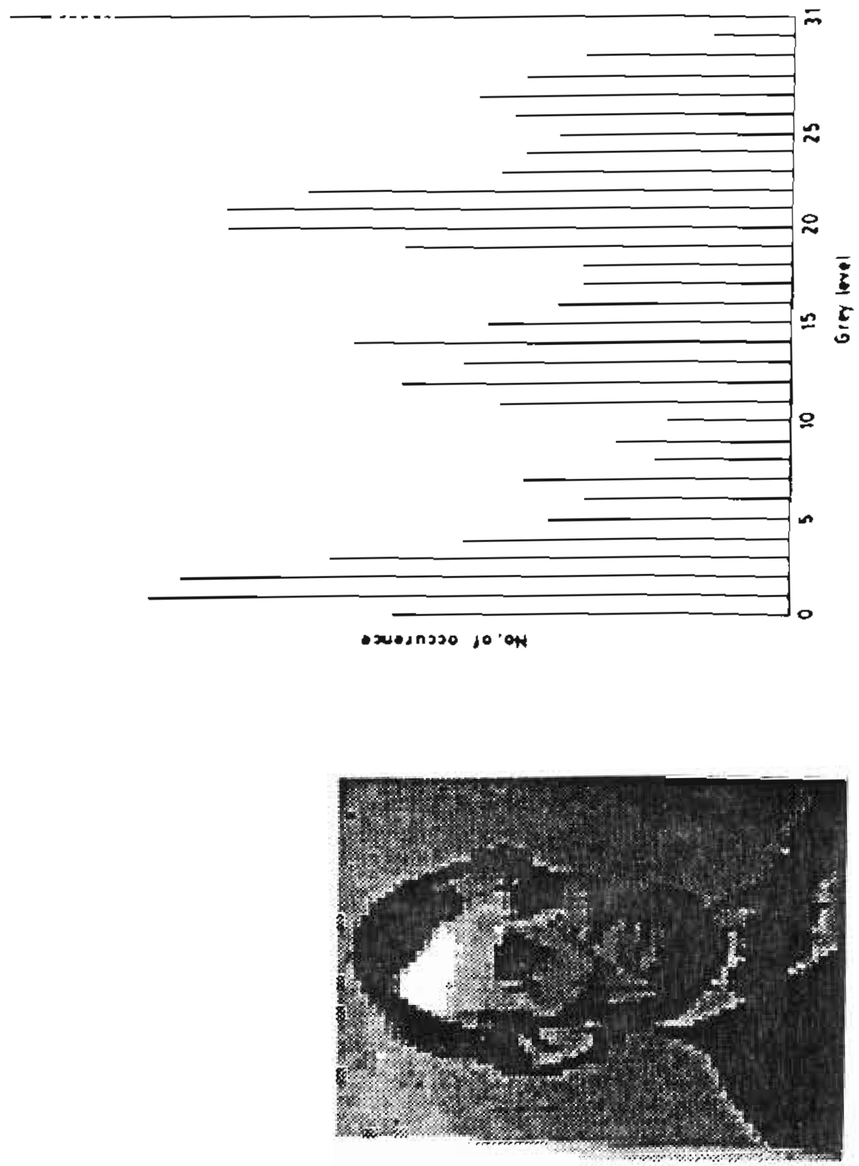


Fig. 7. Bound functions.



(a) (b)

Fig. 8. a) Biplane image; b) its histogram.



(a) (b)

Fig. 9. a) Lincoln image; b) its histogram

TABLE 1

Thresholds for Lincoln Image Using Equation (6) and Entropy

Window Size c	Values for k (1)	(2)	(3)	(4)	(5)	(6)
4	10, 18, 25	7, 11, 19, 26	7, 11, 19, 26	7, 11, 14, 19, 26	7, 9, 11, 14, 18, 26	8, 10, 12, 15, 19, 27
5	10, 17, 25	11, 19, 26	7, 11, 19, 26	7, 11, 19, 26	7, 11, 19, 26	7, 11, 14, 19, 26
6	10, 18, 25	10, 18, 25	11, 19, 26	11, 19, 26	7, 11, 19, 26	7, 11, 19, 26
7	9, 17	11, 18, 25	11, 19, 26	11, 19, 26	11, 19, 26	11, 19, 26
8	9, 17	11, 18, 26	10, 18, 25	11, 19, 26	11, 19, 26	11, 19, 26
9	9, 16	10, 18, 24	10, 18, 25	11, 19, 26	11, 19, 26	11, 19, 26
10	9	10, 17, 24	11, 18, 25	10, 18, 25	11, 19, 26	11, 19, 26

TABLE 2

Thresholds for Lincoln Image Using Equation (7) and Entropy

Window Size c	Values for k (1)	(2)	(3)	(4)	(5)	(6)
4	10, 18, 25	9, 17, 24	6, 10, 14, 18, 24	6, 9, 11, 14, 18, 25	6, 9, 11, 14, 18, 25	5, 8, 10, 13, 17, 24
5	10, 17, 25	9, 17, 24	9, 17, 24	6, 9, 14, 17, 24	6, 9, 11, 14, 18, 25	6, 9, 11, 14, 18, 25
6	10, 18, 25	10, 17, 25	9, 17, 24	9, 17, 24	6, 9, 14, 24, 17	6, 9, 11, 14, 18, 24
7	9, 17	10, 17, 25	9, 16, 24	9, 17, 24	9, 17, 24	9, 14, 17, 24
8	9, 17	9, 17, 25	10, 17, 25	9, 16, 24	9, 17, 24	9, 17, 24
9	9, 16	9, 16	10, 17, 25	9, 16, 24	9, 17, 24	9, 17, 24
10	9	8, 16	9, 17, 25	10, 17 25	9, 16 24	9, 17, 24

TABLE 3

Thresholds for Lincoln Image Using Equation (6) and Index of Fuzziness

Window Size c	Values for k (1)	(2)	(3)	(4)	(5)	(6)
4	11, 18, 25	7, 11, 19, 26	7, 9, 11, 14, 18, 26	7, 9, 11, 14 18, 26	7, 9, 11, 14, 18, 26	8, 10, 12, 15, 19, 26
5	10, 17, 25	11, 19, 26	7, 11, 19, 26	7, 11, 14, 19, 26	7, 9, 11, 14, 18, 26	7, 9, 11, 14, 18, 26
6	10, 18, 25	10, 18, 25	7, 11, 19, 26	7, 11, 19, 26	7, 11, 14, 19, 26	7, 9, 11, 14, 18, 26
7	9, 17, 25	11, 18, 25	11, 19, 26	11, 19, 26	11, 19, 26	11, 19, 26
8	10, 17, 26	11, 18, 26	10, 18, 25	11, 19, 26	11, 19, 26	11, 19, 26
9	9, 17	10, 18, 25	11, 18, 25	11, 19, 26	11, 19, 26	11, 19, 26
10	10, 17	10, 18, 25	11, 18, 25	10, 18, 25	11, 19, 26	11, 19, 26

TABLE 4

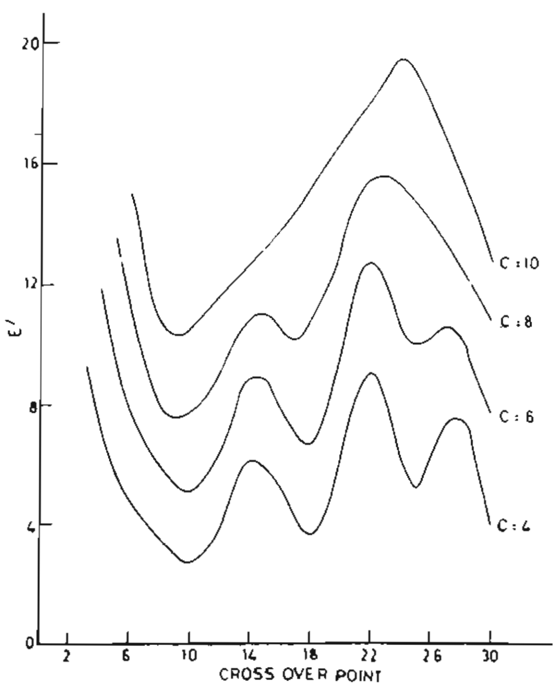
Thresholds for Biplane Image Using Equation (6) and Entropy

Window Size c	Values for k (1)	(2)	(3)	(4)	(5)	(6)
5	14, 23	15, 24	13, 15, 24	13, 15, 24	12, 15, 24	12, 15, 24
6	13	14, 23	15, 24	13, 15, 24	13, 15, 24	12, 15, 24
7	13	14	15, 24	15, 24	13, 15, 24	13, 15, 24
8	13	15	14, 23	15, 24	15, 24	13, 15, 24
9	13	14	14	15, 24	15, 24	15, 24
10	13	14	15	14, 23	15, 24	15, 24
11	13	15	15	15	15, 24	15, 24
12	14	14	15	15	15, 23	15, 24

TABLE 5

Thresholds for Biplane Image Using Equation (7) and Entropy

Window Size c	Values for k (1)	(2)	(3)	(4)	(5)	(6)
5	13	12, 22	12, 22	12, 22	12, 22	12, 18, 22
6	13	12	12, 22	12, 22	12, 22	12, 22
7	13	12	11	12, 22	12, 22	12, 22
8	13	12	12	11	12, 22	12, 22
9	12	13	12	11	11, 22	12, 22
10	13	12	12	12	11	11, 22
11	13	12	12	12	11	11
12	14	12	11	12	12	11



Figs. 10–12. Entropy values for Lincoln image using Equation (6) are plotted against crossover points for $K = 1, 2, 3$. $E' = [E \times MN \ln 2]/20$.

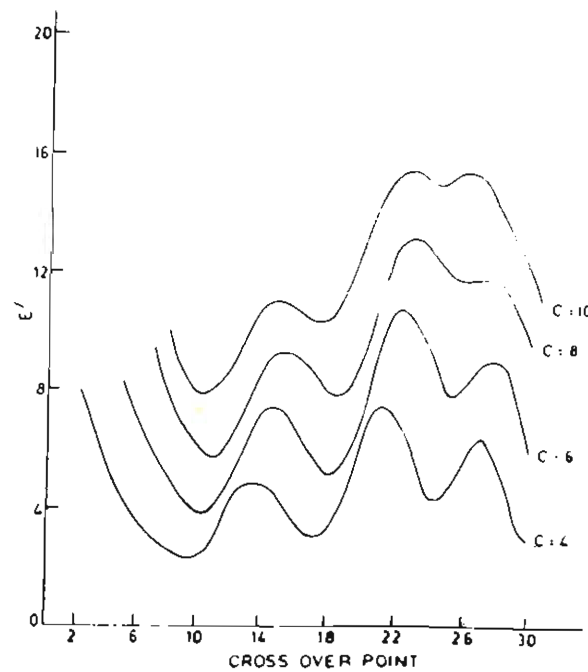
Note 3.

Let g be a function (Figure 6) from $[0, c]$ to $[0, 1]$ such that

(i) $g(0) = 0$, $g(c) = 1$, g is monotonically nondecreasing.
(ii) There exists a point $z \in [3c/4, c)$ such that $g(z) = 1/2$.
(iii) There exists z_0, $c/2 \leqslant z_0 < z$ such that $g(z_0) < \epsilon$, where ϵ is a small positive value.

That is, most of the variation in g is concentrated towards the end point c. The multiplication of the heights of the histogram with either Shannon's function or index of fuzziness would be insignificant in the interval $[0, c/2]$ if ϵ is taken suitably. This would essentially result in a membership function g whose domain is of length $c/2$ but not c.

That means that, once the value of c is chosen, the variation of q should not be concentrated mostly on a small interval towards the end point of the interval $[0, c]$. Similarly, it can be argued that it cannot be concentrated towards the starting point of the interval $[0, c]$. From Note 2, it is apparent

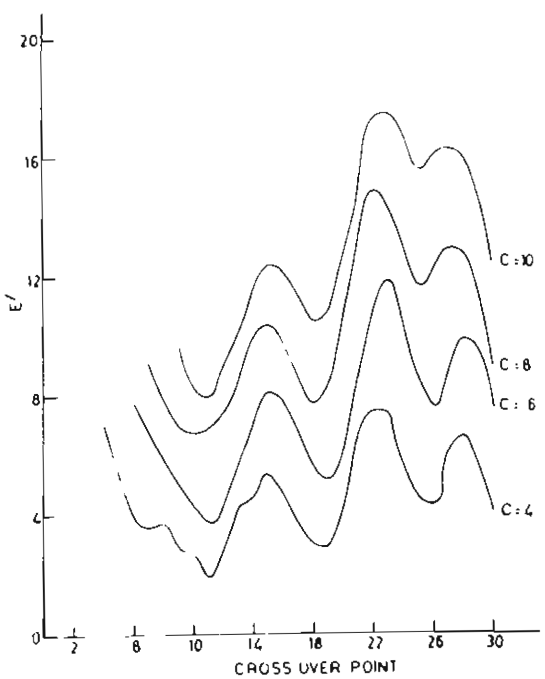


that the variation in g cannot be concentrated in a small interval in the middle of $[0, c]$.

Observe that if the variation in g is concentrated towards one of the end points of g, or in a small interval in the middle of $[0, c]$, then it cannot satisfy the bounds of Murthy and Pal [8]. The conclusion is that g can be taken to be a function satisfying the above bounds. In Figure 7, such bound functions are shown. Once g satisfies the bounds, then the form of g may be taken as $g(x) = 1 - g(c - x)\ \forall x \in [0, c]$, because this would give the valley, in case it is present.

5. EXPERIMENTAL RESULTS

To verify the results, two images are considered. They are (i) biplane (Figure 8) and (ii) Lincoln (Figure 9) along with their histograms. Two types of



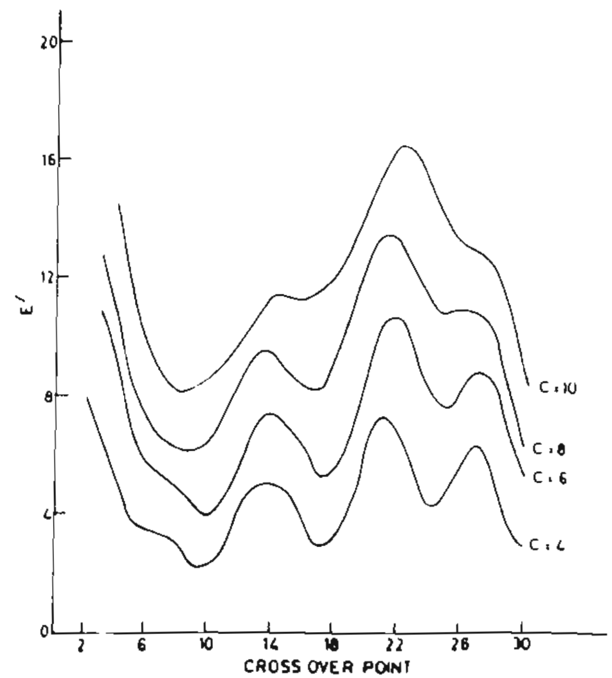
membership functions are considered:

$$\text{(i)}\quad g_{1k}(x) = (x/c)^k, \qquad 0 \leqslant x \leqslant c, \quad k \text{ is a positive integer}, \tag{6}$$

$$\text{(ii)}\quad g_{2k}(x) = 1 - (1 - x/c)^k, \qquad 0 \leqslant x \leqslant c, \quad k \text{ is a positive integer}. \tag{7}$$

Observe that most of the variation in g_{1k} will be concentrated towards c as k increases. Similarly, most of the variation in g_{2k} will be concentrated towards 0 as k increases. The relation between this sort of functions and thresholds is discussed in Note 3 of Section 4. g_{1k} and g_{2k} are specifically selected for this reason.

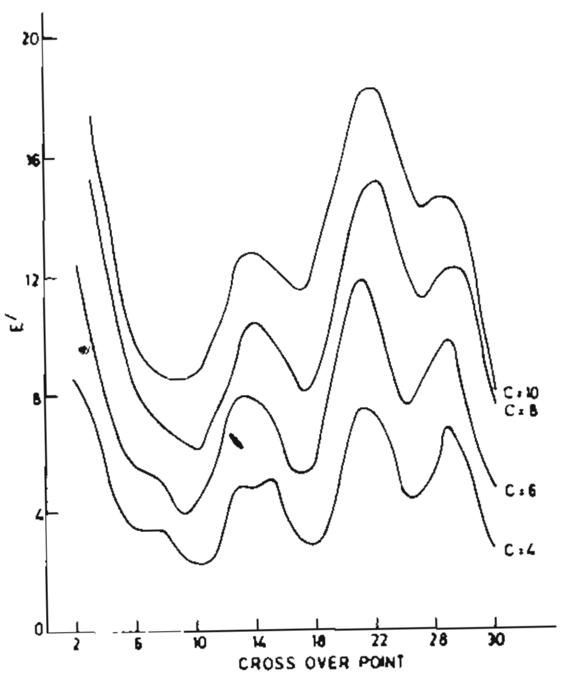
The values of k for both g_{1k} and g_{2k} are varied from 1 to 6. Simultaneously the window size c is also varied. For a fixed c and k, the window is moved throughout the dynamic range of the grey level. The corresponding



Figs. 13 and 14. Entropy values for Lincoln image using Equation (7) are plotted against crossover points for $K = 2, 3$.

fuzzy measures are computed. The crossover points of membership function for which the measures take local minimum values are found out. They are, as explained in Section 2, taken to be the optimum thresholds for crisp segmentation. The threshold values thus obtained are shown in Tables 1–5. These threshold values are nothing but the crossover points of the membership function for that particular position of the window. Tables 1, 2, 4, and 5 use entropy as a measure of greyness ambiguity whereas Table 3 uses index of fuzziness. The similarities between index of fuzziness and entropy can be observed from Tables 1 and 3. The results agree well with the mathematical conclusions arrived at in Sections 3 and 4.

The window size c is a function of the distance between modes between which a valley is to be detected. It is seen from the above-mentioned tables that as c increases, some of its valleys disappear for a fixed k. For example, the number of valleys for Lincoln image reduces (Table 1) from three to one as c increases from 4 to 10 for $k = 1$. Figures 10–14 illustrate the said fact graphically. As a typical illustration, this is shown only for Lincoln image



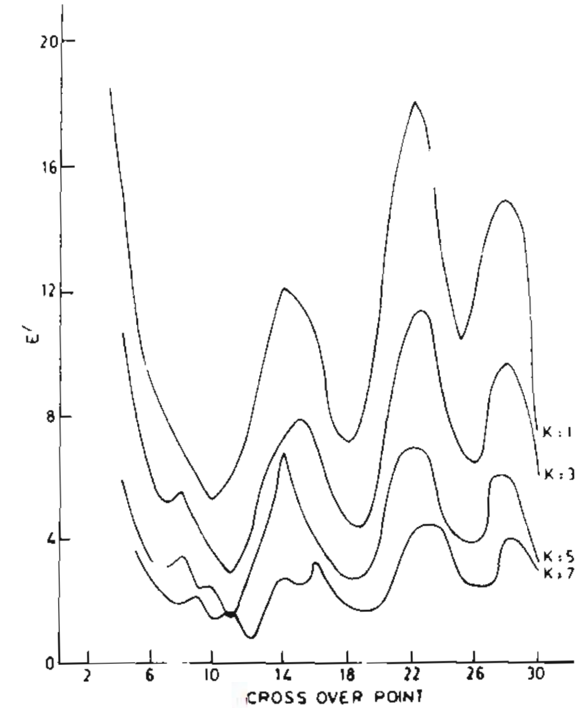


Fig. 15. Entropy values for Lincoln image using Equation (6) are plotted against crossover points for fixed $c = 4$.

(Figure 9). Figures 10–12 correspond to Equation (6) for $k = 1, 2, 3$ and Figures 13 and 14 correspond to Equation (7) for $k = 2, 3$. These findings corroborated an earlier result [5] for a 256 level multimodal x-ray image using Zadeh's S function [Equation (4)].

For a given window size c, the number of valleys increases giving rise to redundant thresholds as k increases. Observe that, as k increases, most of the variation in membership function gets shifted to one of the end points of the window for both g_{1k} and g_{2k}. This agrees with the mathematical justification given in Section 4. This is further demonstrated in Figures 15 and 16 for $c = 4$ and Lincoln image as input. Table 6 gives the result of using a specific class of membership functions which have symmetry in greyness ambiguity. This table is discussed in the next section elaborately where optimal choices of c and g are explained in relation to the images of biplane and Lincoln.

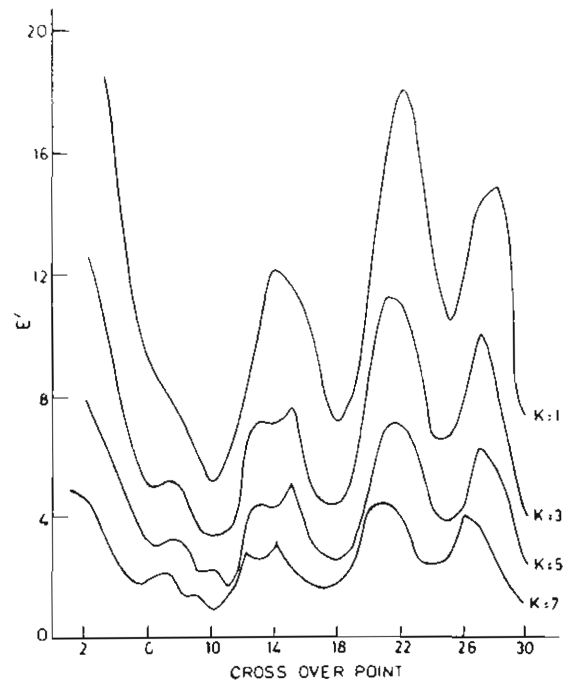


Fig 16. Entropy values for Lincoln image using Equation (7) are plotted against crossover points for fixed $c = 4$.

TABLE 6
Thresholds for Lincoln Image Using Equation (9) and Entropy

	Values of k								
c	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
4	10, 18, 25	10, 18, 25	10, 18, 25	11, 18, 25	11, 18, 25	7, 11, 18, 25	7, 11, 18, 25, 14	7, 11, 18, 25, 14	7, 9, 14, 18, 11, 25
5	10, 18, 25	10, 18, 25	10, 18, 25	10, 18, 25	11, 18, 25	11, 18, 25	11, 18, 25	7, 11, 18, 25	7, 11, 14, 18, 25

6. DISCUSSION ON OPTIMAL CHOICE AND CONCLUSIONS

The experimental results of Section 5 are discussed here elaborately in relation to the mathematical setup of Sections 3 and 4.

Note that the biplane image has two modes whereas Lincoln has four modes in the same dynamic range of 32 grey levels. Therefore, the definition of sharp peaks is comparatively difficult in case of Lincoln image. It is also to be noted that the theory in Section 3 demands the histogram to possess the properties of continuity, convexity, and symmetry. In practice, in a digital image, the continuity property is lost. Besides this, it is quite likely that the histogram of a picture may not have the convexity and symmetry properties. The experimental results are therefore to be analyzed, taking these limitations into account.

Let us now consider the image of biplane where the theory demands the value of c be less than 21 (difference between the two dominant modes) in order to detect one valley. Observe that neither g_{1k} nor g_{2k} satisfies the bounds [8] for $k \geqslant 3$. So let us concentrate over the results for $k < 3$ only. From Table 4, it is seen that the single threshold is obtained even if one considers a small value of $c = 7$ for $k = 1, 2$. If c is reduced further (i.e., too small), the algorithm is seen to be able to, according to the theory, detect another threshold between the intermediate modes 21 and 27. It is because the difference between them is 6. It therefore appears that the histogram can be sharpened using g_{1k} in two parts even for a smaller value of c such as 7. Similar argument and results hold good for Table 5.

It is further to be noted that the functions g_{1k} and g_{2k} for $k > 1$ do not satisfy the following property assumed in Theorem 1:

$$g(x) = 1 - g(c - x), \quad x \in [0, c]. \tag{8}$$

The importance of this condition is better reflected in the case of Lincoln image.

For Lincoln image, the theory demands the value of c be less than 5 (or 6) in order to detect three valleys in between four modes. Since there are four regions, in contrary to only two of biplane, over the same dynamic range, the selection of c is therefore comparatively crucial, because its value is very small. Here, also, since g_{1k} and g_{2k} don't satisfy the bounds [8] for $k \geqslant 3$, we concentrate on $k < 3$. Considering Table 1, it is seen that for $c \geqslant 7$ and for $k = 1$, one of the desirable thresholds, as expected, disappears.

The importance of the Equation (8) can be observed from Tables 1 and 3, where, for $c = 4$, both the tables show four thresholds for g_{1k}, where $k = 2$. In this connection, observe that $g_{1k} = g_{2k}$ for $k = 1$ and it satisfies Equation (8).

g_{1k} and g_{2k} for $k > 1$ do not satisfy Equation (8), but Equation (8) was assumed in Theorem 1. Intuitively the membership function which represents ambiguity in an interval of length c should be such that in the middle of the interval ambiguity should be maximum [i.e., $g(x)$ takes values around 0.5] and ambiguity should decrease uniformly as x moves away from the middle. In fact, Equation (8) aptly describes the above-mentioned idea.

Let us consider the membership function

$$g_k(x) = \frac{2^k}{2}\left(\frac{x}{c}\right)^k, \qquad 0 \leqslant x \leqslant \frac{c}{2}$$

$$= 1 - \frac{2^k}{2}\left(\frac{c-x}{c}\right)^k, \qquad c/2 \leqslant x \leqslant c, \tag{9}$$

which satisfies Equation (8). For $k = 2$, Equation (9) reduces to that of Equation (4). Equation (9) can therefore be viewed as a generalization of Zadeh's S function. We applied $g_k(x)$ on Lincoln image for $c = 4$ and 5 and for different values of k (Table 6). It is found out that the number of thresholds exceeded three when g_k is not within the bounds. Therefore, it may be stated that the membership function has to satisfy not only Equation (8) but also bounds.

In spite of the above findings, observe that the desired thresholds are also obtained for higher values of both c and k as shown in Tables 1–3. The reason for this phenomenon is explained in Section 4. Here, k being higher implies that the variation of the membership function will be essentially in a smaller interval of $[0,c]$ which effectively results in smaller window size. In that smaller interval, the given membership function for higher k may be approximated by a function with smaller k. In the case of Lincoln image, as explained earlier in this section, $c = 4$ or 5 and $k = 1$ are optimal.

In conclusion, minimizing greyness ambiguity by different fuzzy measures can be used to formulate a method of segmentation or threshold detection or sharpening of a grey tone image. The results obtained by any kind of monotonically nondecreasing membership function satisfying the bounds and Equation (8) are seen to be acceptable for the above-mentioned purposes. (For example, for the biplane image, the thresholds as expected were found to be detected anywhere over the flat valley.) In this sense, the proposed fuzzy set theoretic approach is flexible but effective.

It is to be mentioned here that the ambiguity measures considered in this paper are entropy and index of fuzziness. Entropy uses Shannon's function whereas the index of fuzziness uses linear distance between a fuzzy set and its nearest ordinary set. There are other measures which behave similarly as far as greyness ambiguity is concerned. They are, namely, the quadratic index of fuzziness [2], which measures the quadratic distance between a fuzzy set and its nearest ordinary set, and a new entropy measure [9], which was introduced recently. This new entropy incorporates exponential behavior of information gain instead of logarithmic behavior as in the case of Shannon's entropy. Similar mathematical results can also be obtained for them.

7. SUMMARY

The problem of histogram sharpening and thresholding by minimizing ambiguity in greyness using the measures of fuzziness in a set is considered. The earlier works in this context for obtaining both fuzzy and crisp segmentation and their incompleteness are discussed. The earlier works were essentially experimental and no mathematical basis to the results obtained was provided. The criteria regarding the choice of membership functions were not provided in them. Also the choice of the size of the window was determined only experimentally but not established mathematically.

This paper provides a complete mathematical formulation of the above problem. A functional form for the membership function has been given which reflects the intuitive idea behind greyness ambiguity [Equation (8)]. This functional form gives symmetry in greyness ambiguity around the crossover point. It has been established mathematically that a valley in histogram will be obtained if the window size is less than the difference between the modes. Various membership functions are considered, and it has been concluded mathematically that a membership function lying in between the bound functions with symmetry in ambiguity gives the desired valleys. This method is flexible enough and provides effective results. Experimental results on various images further establish the same conclusions.

APPENDIX: BOUNDS FOR MEMBERSHIP FUNCTIONS

The membership function g considered in this paper throughout has the following properties:

(i) $g: [0,c] \to [0,1]$ is continuous,
(ii) $g(0) = 0$, $g(1) = 1$ and g is monotonic.

Recently, Murthy and Pal [8] formulated bounds for membership functions of the above sort in order to discard the membership functional forms which are to be avoided while representing a fuzzy set in practice. Significance of these bounds in image segmentation and analysis problems was also found to be justified [8].

The expression for bound functions are based on the properties of correlation [10] between two membership functions $\sigma_1(x)$ and $\sigma_2(x)$. The main properties on which correlation was formulated are

P_1: If, for higher values of σ_1, σ_2 takes higher values and, for lower values of σ_1, σ_2 also takes lower values, then $c_{\sigma_1,\sigma_2} > 0$ (c represents correlation).

P_2: If $\sigma_1 \uparrow$ and $\sigma_2 \uparrow$, then $c_{\sigma_1,\sigma_2} > 0$.

P_3: If $\sigma_1 \uparrow$ and $\sigma_2 \downarrow$ then $c_{\sigma_1,\sigma_2} < 0$ ($\uparrow$ denotes increases and $\downarrow$ denotes decreases).

It is to be mentioned that P_2 and P_3 should not be considered in isolation of P_1. Had this been the case, one can cite several examples when $\sigma_1 \uparrow$ and $\sigma_2 \uparrow$ but $c_{\sigma_1,\sigma_2} < 0$ and $\sigma_1 \uparrow$ and $\sigma_2 \downarrow$ but $c_{\sigma_1,\sigma_2} > 0$. Subsequently, the type of membership functions which should not be considered in fuzzy set theory are categorized with the help of correlation. Bound functions h_1 and h_2 are accordingly derived [8]. They are

$$h_1(x) = 0, \qquad 0 \leqslant x \leqslant \epsilon$$
$$= x - \epsilon, \qquad \epsilon \leqslant x \leqslant 1,$$
$$h_2(x) = x + \epsilon, \qquad 0 \leqslant x \leqslant 1 - \epsilon$$
$$= 1, \qquad 1 - \epsilon \leqslant x \leqslant 1,$$

where $\epsilon = 0.25$. The bounds for membership function g considered throughout this paper are $h_1(x) \leqslant g(x) \leqslant h_2(x)$ for $x \in [0, 1]$.

For x belonging to any arbitrary interval, the bound functions will be changed proportionately. For $h_1 \leqslant g \leqslant h_2$, $c_{h_1,h_2} \geqslant 0$, $c_{h_1,g} \geqslant 0$, and $c_{h_2,g} \geqslant 0$. The function g lying in between h_1 and h_2 does not have most of its variation concentrated (i) in a very small interval, (ii) towards one of the end points of the interval under consideration, and (iii) towards both the end points of the interval under consideration.

The authors gratefully acknowledge Professor D. Dutta Majumder for his advices and discussions during the course of work. They are also grateful to Mr. S. Chakraborty for drawing diagrams, Mr. Tapas Das for typing the manuscript, and Mr. Ashish Ghosh for writing the computer programs.

REFERENCES

1. S. K. Pal and D. Dutta Majumder, *Fuzzy Mathematical approach to Pattern Recognition*, Wiley (Halsted Press), New York, 1986.
2. A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets—Fundamental Theoretical Elements*, Vol. 1, Academic, New York, 1975.

3. A. Deluca and S. Termini, A definition of a non-probabilistic entropy in the setting of fuzzy set theory, *Inform. Control* 20:301–312 (1972).
4. S. K. Pal, A measure of edge ambiguity using fuzzy sets, *Pattern Recognition Lett.* 4:51–56 (1986).
5. S. K. Pal, R. A. King, and A. A. Hashim, Automatic greylevel thresholding through index of fuzziness and entropy, *Pattern Recognition Lett.* 1:141–146 (1983).
6. S. K. Pal and A. Rosenfeld, Image enhancement and thresholding by optimization of fuzzy compactness, *Pattern Recognition Lett.* 7:77–86 (1988).
7. L. A. Zadeh, Calculus of fuzzy restrictions, in *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, (L. A. Zadeh et al., Eds.) Academic, London, 1975, pp. 1–39.
8. C. A. Murthy and S. K. Pal, Bounds for membership function: Correlation based approach, *Fuzzy Sets and Systems*: (1989), Communicated.
9. N. R. Pal and S. K. Pal, Object background segmentations using new definitions of entropy, *IEE Proc. Part E*, 136(4):284–295 (1989).
10. C. A. Murthy, S. K. Pal, and D. Dutta Majumder, Correlation between two fuzzy membership functions, *Fuzzy Sets Syst.* 17(1):23–38 (1985).

Received 10 March 1989; revised 30 June 1989