Fuzzy divergence, probability measure of fuzzy events and image thresholding

Dinabandhu Bhandari, Nikhil R. Pal* and D. Dutta Majumder

Electronics and Communication Sciences Unit, Indian Statistical Institute, 203 B.T. Road, Calcutta 700 035, India

Received 12 August 1991
Revised 16 April 1992

Abstract

Bhandari, D., N.R. Pal and D. Dutta Majumder, Fuzzy divergence, probability measure of fuzzy events and image thresholding, Pattern Recognition Letters 13 (1992) 857–867.

A new measure called divergence between two fuzzy sets is introduced along with a few properties. Its application to clustering problems is indicated and applied to an object extraction problem. A tailored version of the probability measure of a fuzzy event is also used for image segmentation. Both parametric and non-parametric probability distributions are considered in this regard.

Keywords. Fuzzy divergence, fuzzy event, fuzzy dissimilarity, segmentation.

1. Introduction

This paper is logically divided into two parts. In the first, a new measure called divergence between two fuzzy sets (fuzzy divergence) is introduced. In classical probability space a measure of divergence exists, which quantifies the discrepancy between two probability distributions. In this note we introduce the concept of fuzzy divergence which represents a measure of dissimilarity or difference between two fuzzy sets. It may be mentioned that the divergence measure is not a metric. A few propositions about the measure are also made. As an illustration of its applicability, it has been used in a clustering problem for object–background classification.

In the second part, the probability measure of fuzzy event introduced by Zadeh (1986) is reviewed. It has been found that a tailored version of this probability measure can be used as a quantitative index of similarity/dissimilarity between two sets. Hence, this can also be used for clustering/segmentation problems. In addition, algorithms have been developed for image thresholding using this measure. Both parametric and non-parametric distributions for describing the histogram have been explored. Under the parametric approach, the normal distribution as well as the Poisson distribution have been considered.

For the parametric methods initially, a coarse object–background classification is carried out, minimizing the χ^2 statistic. The parameters of the

Correspondence to: D. Bhandari, Electronics and Communication Sciences Unit, Indian Statistical Institute, 203 B.T. Road, Calcutta 700 035, India.

* Currently with the Division of Computer Science, Univ. of West Florida, 11000 University Parkway, Pensacola, FL 32514, USA.

resultant distribution are then used to select the membership function for defining the required fuzzy set. Finally, the fuzzy dissimilarity measure between object and background is maximized to select the threshold for segmentation. The superiority of the proposed algorithms has been established by comparing the results with some existing methods.

2. Divergence measure for fuzzy sets and its application

2.1. Divergence between two fuzzy sets

A crisp subset A of the universal set U is a collection of objects from U, which are members of A. An equivalent way of defining A is to specify the characteristic function of A, $\chi_A: U \to \{0,1\}$ for all $x \in U$, such that

$$\chi_A(x) = 1, \quad x \in A,$$
$$= 0, \quad x \notin A.$$

Generalizing the characteristic function from $\{0,1\}$ to $[0,1]$ one can obtain the fuzzy sets. More specifically, the above concept of characteristic function generalizes to a membership function $\mu: U \to [0,1]$. In general a fuzzy set A in the universe of discourse is defined as

$$A = \{\mu_A(x_i) \mid x_i,\ i = 1,2,\ldots,n\},$$

where $\mu_A(x_i)$ is the membership value for x_i indicating the degree of possessing the property A.

Let S be the set of supports x_i, $i=1,2,\ldots,n$, and $A = \{\mu_1(x_i) \mid x_i\}$ and $B = \{\mu_2(x_i) \mid x_i\}$ be two fuzzy sets defined on S. Following the concept of divergence in classical probability theory (Kullback (1959)), we define the divergence $D(A,B)$ between A and B as

$$D(A,B) = \frac{1}{n}\sum_{i=1}^{n}[D_i(A,B) + D_i(B,A)] \qquad (1)$$

where

$$D_i(A,B) = \mu_1(x_i)\log\frac{\mu_1(x_i)}{\mu_2(x_i)} + [1-\mu_1(x_i)]\log\frac{1-\mu_1(x_i)}{1-\mu_2(x_i)}$$
$$0<\mu_1(x_i),\ \mu_2(x_i)<1 \qquad (2)$$

and

$$D_i(B,A) = \mu_2(x_i)\log\frac{\mu_2(x_i)}{\mu_1(x_i)} + [1-\mu_2(x_i)]\log\frac{1-\mu_2(x_i)}{1-\mu_1(x_i)}$$
$$= -\mu_2(x_i)\log\frac{\mu_1(x_i)}{\mu_2(x_i)} - [1-\mu_2(x_i)]\log\frac{1-\mu_1(x_i)}{1-\mu_2(x_i)}$$
$$0<\mu_1(x_i),\ \mu_2(x_i)<1. \qquad (3)$$

Here, $\sum_{i=1}^{n} D_i(A,B)$ can be described as the mean information per support from A for discrimination in favor of A against B. A similar interpretation is also applicable for $\sum_{i=1}^{n} D_i(B,A)$. The second part in D_i's has been incorporated to bring into account the fact that the divergence between the complements of A and B should be equal to that between A and B.

Note that $D(A,B)$ is symmetric with respect to A and B, and it has all the metric properties except the triangle inequality property.

Property 1. $D(A,B) \geqslant 0$, $D(A,B)=0$ *iff* $A=B$.

Property 2. $D(A,B) = D(B,A)$.

It is also interesting to note the following proposition.

Proposition 1. *For any two fuzzy sets A and B,*

$$D(A\cup B, A\cap B) = D(A,B).$$

This is indeed a desirable property for any distance measure between two fuzzy sets.

It is to be mentioned here that equations (2) and (3) do not include the crisp sets. In order to account for this, one can use the following expressions for D_i's:

$$D_i(A,B) = \mu_1(x_i)\log\frac{1+\mu_1(x_i)}{1+\mu_2(x_i)} + [1-\mu_1(x_i)]\log\frac{2-\mu_1(x_i)}{2-\mu_2(x_i)}$$

and

$$D_i(B,A) = \mu_2(x_i)\log\frac{1+\mu_2(x_i)}{1+\mu_1(x_i)}$$

$$+[1-\mu_2(x_i)]\log\frac{2-\mu_2(x_i)}{2-\mu_1(x_i)}$$

$$0\leqslant\mu_1(x_i),\ \mu_2(x_i)\leqslant 1. \tag{4}$$

This does not violate the properties satisfied by $D(A,B)$ defined earlier. Under this framework the following propositions can be stated.

Let $A=\{\mu_A(x_i)\mid x_i\}$ be a fuzzy set. The furthest non-fuzzy set $\bar{A}$ is defined as $\bar{A}=\{\mu_{\bar{A}}(x_i)\mid x_i\}$, where

$$\mu_{\bar{A}}(x_i) = 1 \quad \text{if } \mu_A(x_i)\leqslant 0.5,$$
$$= 0 \quad \text{otherwise.}$$

Proposition 2. *For any fuzzy set A, $D(A,B)$ is maximum iff B is the furthest non-fuzzy set ($\bar{A}$) of A. In other words,*

$$\max_B D(A,B) = D(A,\bar{A}).$$

Proposition 3. *Let A^c be the complement of A, then*

$$\max_A D(A,A^c) = 2\log 2.$$

This occurs when A is a non-fuzzy set.

2.2. Applications

The divergence measure introduced in the previous section can be used for clustering problems as this information measure may be used to quantify the separation between classes. Here, as an illustration of its applicability we shall use it for image segmentation.

A grey tone $M\times N$ image of L levels can be considered as an array of fuzzy singletones, each having a value of membership denoting its degree of belonging to object (black) and background (white) relative to some brightness level x_i: $x_i=0,1,2,\ldots,L-1$. Let $\mu_o(x_i)$ and $\mu_b(x_i)$ be the degree of belonging of an image pixel having grey level x_i $(0\leqslant x_i\leqslant L-1)$ to the object O and to the background B, respectively. In other words, $\{\mu_o(x_i)\mid x_i\}$ represents the fuzzy set "*X is black (object)*" while $\{m_b(x_i)\mid x_i\}$ characterizes the fuzzy set "*X is white (background)*". According to (1) and (4) the divergence between object and background of the image can be defined as

$$D(O,B)=\frac{1}{MN}\sum_{i=0}^{L-1}h(x_i)[D_i(O,B)+D_i(B,O)] \tag{5}$$

where $h(x_i)$ is the frequency (number of occurrences) of grey level x_i in the image,

$$D_i(O,B)=\mu_o(x_i)\log\frac{1+\mu_o(x_i)}{1+\mu_b(x_i)}+[1-\mu_o(x_i)]\log\frac{2-\mu_o(x_i)}{2-\mu_b(x_i)}$$

and

$$D_i(B,O)=\mu_b(x_i)\log\frac{1+\mu_b(x_i)}{1+\mu_o(x_i)}+[1-\mu_b(x_i)]\log\frac{2-\mu_b(x_i)}{2-\mu_o(x_i)}$$

$$x_i=0,1,2,\ldots,L-1. \tag{6}$$

In other words,

$$D(O,B)=\frac{1}{MN}\sum_{x_i=0}^{L-1}h(x_i)\times\left\{[\mu_o(x_i)-\mu_b(x_i)]\log\frac{1+\mu_o(x_i)}{1+\mu_b(x_i)}+[\mu_b(x_i)-\mu_o(x_i)]\log\frac{2-\mu_o(x_i)}{2-\mu_b(x_i)}\right\}. \tag{7}$$

Let us now assume that s, $0<s<L-1$, is a threshold for object-background classification of the image F and that $D(O,B:s)$ is the divergence measure corresponding to s. In this sense, s is the most ambiguous point on the grey scale. Thus the membership functions are to be chosen in such a manner that $\mu_o(s)=\mu_b(s)=0.5$ (Figure 1) and increase as we go away from s. Any S-type function with the above requirement can be used. However, a detailed discussion on the selection of a membership function can be found in Section 4.

It is clear that the divergence measure should be maximum when s corresponds to an appropriate valley for object-background classification. Thus the optimum threshold can be obtained by maximizing $D(O,B:s)$ with respect to s. In other words, τ can be taken as the threshold, when

$$D(O,B:\tau)=\max_s D(O,B:s). \tag{8}$$

Before describing the results obtained by the method, another algorithm using the probability measure of fuzzy events will be developed in the following section.

3. Probability measure of a fuzzy event and its application

3.1. *Probability measure of a fuzzy event*

Let us now define a fuzzy event and its fuzzy probability measure. In ordinary probability theory, a probability space is a triplet (Ω, B, P), where B is the σ-field of Borel sets in Ω and p is a probability measure over Ω, such that $0 \leqslant P(A) \leqslant 1$, for any $A \in B$, with $P(\emptyset)=0$, $P(\Omega)=1$ and it satisfies the countably additivity property, i.e., if $A_1, A_2, \ldots, A_n$ are disjoint events of B then,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} p(A_i).$$

For an event $A \in B$, the probability of A can be expressed as

$$P(A) = \int_A \mathrm{d}P$$

or

$$P(A) = \int_\Omega \chi_A(x)\,\mathrm{d}P = E(\chi_A). \tag{9}$$

Here, χ_A defines the characteristic function of A ($\chi_A(x)=0$ or 1) and $E(\chi_A)$ is the expectation of χ_A.

The notion of an event and its probability constitute the most basic elements of probability theory. As defined above, an event is a precisely specified collection of points in the sample space. By contrast, in real life one frequently encounters situations in which an event is fuzzy rather than a clearly defined collection of points (Kandel (1982)). For example, the ill-defined events: "x is a *tall* man", "x is *much greater* than 1" are fuzzy because of the imprecision in the meaning of the italicized words.

By using the above concept Zadeh (1986) extended the notion of an event and its probability to the fuzzy domain. Let (Ω, B, P) be a probability space in which B is a σ-field of Borel sets in Ω and P is the probability measure over Ω. Then a fuzzy event in Ω is a fuzzy set A in Ω whose membership function $\mu_A : \Omega \to [0,1]$ is Borel measurable. The probability of a fuzzy event A is defined by the following:

$$P(A) = \int_\Omega \mu_A(x)\,\mathrm{d}P = E(\mu_A). \tag{10}$$

Thus, as in the case of a crisp set, the probability of a fuzzy event A is the expectation of its membership function.

Under a proper framework, as we shall see in the next section, this probability measure can be viewed as a similarity/dissimilarity measure between two sets.

3.2. *Fuzzy similarity/dissimilarity measure between two sets*

Let X_1 and X_2 be two sets characterized by some joint probability distribution P. Let us define a fuzzy event A,

$$A = \{x_1, x_2 \mid x_1 \in X_1,\ x_2 \in X_2;\ x_1, x_2 \text{ are similar/dissimilar}\}.$$

The fuzzy set A may be defined by a suitable membership function $\mu_A(x_1, x_2)$, which will give the degree of similarity/dissimilarity. Then the probability of the fuzzy event A is defined by

$$P(A) = \int_{X_1}\int_{X_2} \mu_A(x_1, x_2)\,\mathrm{d}P. \tag{11}$$

This $P(A)$ can be viewed as a measure of similarity/dissimilarity between the two populations. If X_1 and X_2 are two disjoint sets and they are governed by two independent probability distributions, say P_1 and P_2, then one can redefine the similarity/dissimilarity measure as

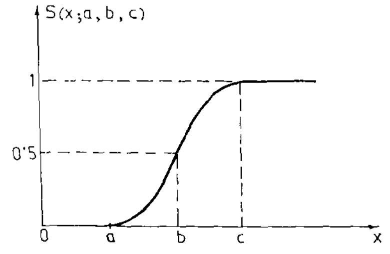


Figure 1. S-function.

$$P(A) = \int_{X_1}\int_{X_2} \mu_A(x_1,x_2)\,dP_1\,dP_2. \tag{12}$$

This will represent a measure of similarity/dissimilarity between two independent distributions. This measure, considering only the set of supports observed in the two sets, can be used in solving clustering the segmentation problems. The following section illustrates its application to object extraction.

3.3. Application

The segmentation problem may be viewed as a partitioning of the image into two non-intersecting dissimilar regions. In other words, given a measure of dissimilarity, our intention is to partition the image in such a manner so as to maximize the dissimilarity between the object and background.

Let s be the assumed threshold for object-background classification. Now each of the two resulting pixel populations can be modeled by some probability density $p_i(g_i,s)$, $i=1,2$ (may be parametric/non-parametric). Given a suitable membership function μ_A for the fuzzy set

$$A = \{x_1,x_2 \mid x_1 \in X_1,\ x_2 \in X_2,\ x_1,x_2 \text{ are dissimilar}\},$$

a dissimilarity measure between two sets (object and background) is obtained by

$$P(A,s) = \int_0^s \int_{s+\varepsilon}^{L-1} \mu_A(g_1,g_2)\,p_1(g_1,s) \times p_2(g_2,s)\,dg_1\,dg_2 \tag{13}$$

where ε is an arbitrary small positive quantity.

For the discrete case, it can be written as

$$P(A,s) = \sum_0^s \sum_{s+1}^{L-1} \mu_A(g_1,g_2)\,p_1(g_1,s)\,p_2(g_2,s). \tag{14}$$

It must be mentioned here that equation (13) or (14) is not exactly the probability of the fuzzy event, "(x_1,x_2) are dissimilar" as the limits of integration (or summation) do not span the entire permissible range. Equation (13) or (14) gives a measure of dissimilarity between two sets which might have been generated from two populations characterized by $p_1(g_1,s)$ and $p_2(g_2,s)$. Note that the overlap area between the two probability distributions has not been considered.

Since in this case, the two probability densities are independent, clearly, $P(A,s)$ is a function of s only. Hence, equation (13) can be regarded as an objective criterion for the correct classification performance. The optimum threshold can therefore, be obtained by maximizing $P(A,s)$, in other words, τ is taken to be the optimum threshold for object-background classification, where

$$P(A,\tau) = \max_s P(A,s). \tag{15}$$

Here $p_i(g_i,s)$ can be parametric or non-parametric as discussed in the next sections.

3.3.1. Non-parametric

In this case, we are considering the histogram itself as the representative of the probability distribution of grey values in the image F. Let $h(g)$ be the frequency of grey value g $(0 \leqslant g \leqslant L-1)$ in F, and let $p_1(g_1,s)$ and $p_2(g_2,s)$ be the probability densities for the two sets, namely, object $(g_1 \leqslant s)$ and background $(g_2 > s)$. Then

$$p_1(g_1,s) = h(g_1) \Big/ \left(\sum_{g_1=0}^{s} h(g_1) \right), \quad 0 \leqslant g_1 \leqslant s, \tag{16}$$

$$p_2(g_2,s) = h(g_2) \Big/ \left(\sum_{g_2=s+1}^{L-1} h(g_2) \right), \quad s+1 \leqslant g_2 \leqslant L-1, \tag{17}$$

and the dissimilarity measure $P(A,s)$ can be written as

$$P(A,s) = \sum_0^s \sum_{s+1}^{L-1} \mu_A(g_1,g_2)\,p_1(g_1,s)\,p_2(g_2,s). \tag{18}$$

3.3.2. Parametric

Usually normal distributions (Kittler and Illingworth (1986), Pal and Bhandari (1992)) are used to describe the grey level variation, but recently it has been established by Pal and Pal (1991) that a grey level distribution over a uniform region can be better approximated by a Poisson distribution. In this study both normal and Poisson distributions have been considered. Let $G_O(\lambda_O(s))$ and $G_B(\lambda_B(s))$ be two Poisson distributions for the object and the background grey levels, respectively. The parameters of the two distributions $\lambda_O(s)$ and $\lambda_B(s)$ can be estimated as

$$\lambda_O(s) = \left(\sum_{g=0}^{s} g\,h(g)\right)\bigg/\left(\sum_{g=0}^{s} h(g)\right) \tag{19}$$

and

$$\lambda_B(s) = \left(\sum_{g=s+1}^{L-1} g\,h(g)\right)\bigg/\left(\sum_{g=s+1}^{L-1} h(g)\right). \tag{20}$$

Hence, the dissimilarity measure $P(A,s)$ between object and background becomes

$$P(A,s) = \sum_{g_1=0}^{s}\sum_{g_2=s+1}^{L-1} \mu_A(g_1,g_2)\frac{(\lambda_1(s))^{g_1}}{g_1!}\,\mathrm{e}^{-\lambda_1(s)} \times \frac{(\lambda_2(s))^{g_2}}{g_2!}\,\mathrm{e}^{-\lambda_2(s)}. \tag{21}$$

On the other hand, if we assume that object and background densities follow normal distributions ($N(m_1,\sigma_1)$ and $N(m_2,\sigma_2)$, respectively), then for an arbitrary threshold s the parameters can be estimated as follows:

$$m_i(s) = \left(\sum_{g=a}^{b} g\,h(g)\right)\bigg/\left(\sum_{g=a}^{b} h(g)\right) \tag{22}$$

and

$$\sigma_i^2(s) = \left(\sum_{g=a}^{b} (g-m_i(s))^2 h(g)\right)\bigg/\left(\sum_{g=a}^{b} h(g)\right) \tag{23}$$

where

$$a = \begin{cases} 0 & \text{for } i=1, \\ s+1 & \text{for } i=2, \end{cases} \tag{24}$$

and

$$b = \begin{cases} s & \text{for } i=1, \\ L-1 & \text{for } i=2. \end{cases} \tag{25}$$

In this situation the measure of dissimilarity between object and background becomes,

$$\begin{aligned} P(A,s) &= \int_0^{s-\varepsilon}\int_{s+\varepsilon}^{L-1} \mu_A(g_1,g_2)\,p_1(g_1,s)\,p_2(g_2,s)\,\mathrm{d}g_1\,\mathrm{d}g_2 \\ &= \frac{1}{\sigma_1(s)\sigma_2(s)}\frac{1}{2\pi}\int_0^{s-\varepsilon}\int_{s+\varepsilon}^{\infty} \mu_A(g_1,g_2) \\ &\quad \times \exp\left(-\frac{1}{2}\left(\frac{g_1-m_1(s)}{\sigma_1(s)}\right)^2\right) \\ &\quad \times \exp\left(-\frac{1}{2}\left(\frac{g_2-m_2(s)}{\sigma_2(s)}\right)^2\right)\mathrm{d}g_1\,\mathrm{d}g_2. \end{aligned} \tag{26}$$

The dissimilarity measure $P(A,s)$, in each case (parametric and non-parametric), is explicitly a function of s only. Maximizing $P(A,s)$ on s we can find the optimal threshold.

To reduce the computation overhead, we have used the χ^2-statistic to find a reasonable range of grey values for the threshold. Here, our intention is not to test the goodness of fit but to find some approximate range in which the threshold lies. For this we minimized

$$\chi_s^2 = \sum_{g=0}^{s}\frac{(O_g-E_1)^2}{E_1} + \sum_{g=s+1}^{L-1}\frac{(O_g-E_2)^2}{E_2} \tag{27}$$

where

$$E_i = p_i(g,s)\,N_i, \quad i=1,2$$

and

$$O_g = h(g), \quad N_1 = \sum_{g=0}^{s} h(g), \quad N_2 = \sum_{g=s+1}^{L-1} h(g).$$

Let

$$\chi_\tau^2 = \min_s \chi_s^2,$$

and $m_1(\tau)$ and $m_2(\tau)$ be the means of the two sets when the threshold is τ. It is then reasonable to assume that the optimal threshold belongs to the interval $[m_1(\tau), m_2(\tau)]$. So instead of computing the dissimilarity for all s ($0 \leqslant s \leqslant L-1$) we can maximize $P(A,s)$ over $m_1(\tau) \leqslant s \leqslant m_2(\tau)$.

4. Selection of membership function

In order to segment an image using the divergence measure, we need to define two fuzzy sets "*g is white*" and "*g is black*". For defining such a pair any S-type function and its complement can be used. We have used here the standard S-function of Zadeh as defined below, for the fuzzy set "*g is white*":

$$\begin{aligned} S(g:a,b,c) &= 0 & g \leqslant a, \\ &= 2\frac{(g-a)^2}{(c-a)^2} & a \leqslant g \leqslant b, \\ &= 1-2\frac{(g-c)^2}{(c-a)^2} & b \leqslant g \leqslant c, \\ &= 1 & g \geqslant c, \end{aligned}$$

where b is the cross-over point and $(c-a)$ is the bandwidth. Thus, we can take $\mu_2(g) = S(g: a, s, c)$ and $\mu_1(g) = 1 - \mu_2(g)$ (since the black set is the complement of white), where s is an assumed threshold. It may be mentioned here that for such an S-function several authors (Pal and Dutta Majumder (1986)) have used different bandwidths (windows) without giving any criteria for the selection of an appropriate window size. For this purpose one can use the guideline provided by Murthy and Pal (1990). In this investigation we have used a bandwidth of 10.

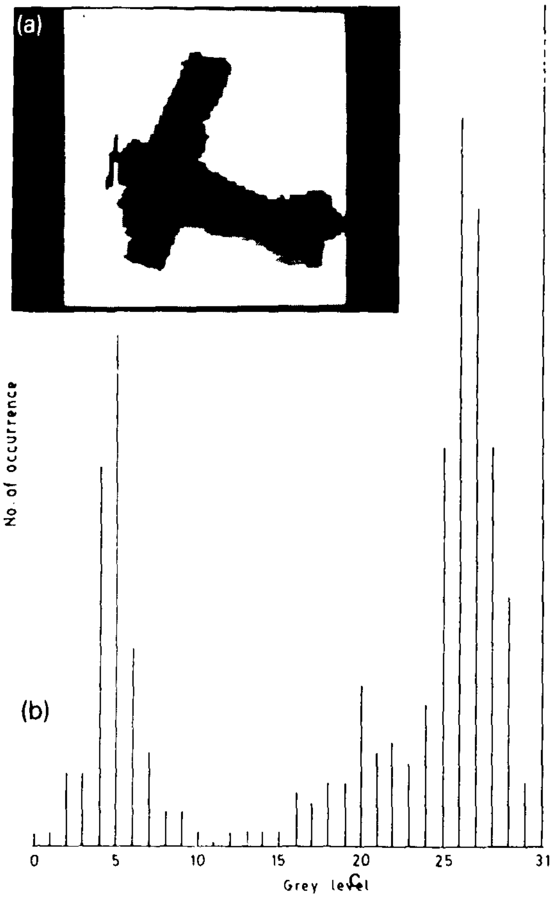
To define the fuzzy set "x_1 *and* x_2 *are dissimilar*", we have used $|x_1 - x_2|$ as the argument of the S-function. Here the window size has been selected depending on the parameters of the probability distributions. For example, one can use $a = 0$ and $c = \lambda_B - \lambda_O$, where λ_O is the parameter of the Poisson distribution assumed for the object and λ_B is that of the background.

Table 1

Images	Thresholds			
	Divergence	Non-param.	Poisson	Normal
Biplane	13	14	12	14
Lincoln	9	11	11	6
Boy	18	14	9	30
Test 1	17	17	17	17
Test 2	10	10	9	9
Test 3	16	16	16	16

One can also use an exponential function for computing the dissimilarity measure between x_1 and x_2. For example,

$$\mu_A(x_1, x_2) = 1 - \exp(-|x_1 - x_2|).$$



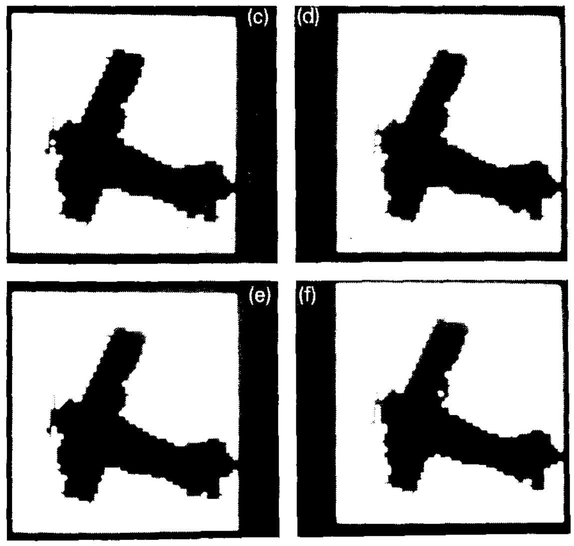


Figure 2. Biplane image. (a) Input. (b) Histogram. (c) Output obtained using divergence. (d) Output obtained using dissimilarity assuming Poisson. (e) Output obtained using dissimilarity assuming normal and non-parametric. (f) Output obtained using minimum error thresholding.

5. Implementation and comparison with some existing methods

In this section, we shall discuss the results obtained by the algorithms introduced earlier on a set of three (64 × 64) images with 32 levels. Algorithms have also been applied on three test histograms. Some of the existing thresholding techniques (Kittler and Illingworth (1986), Pun (1980), Kapur et al. (1985)) have also been implemented and compared with the proposed algorithms.

Kittler and Illingworth (1986) have suggested an iterative method for minimum error thresholding, assuming normal distributions for the grey level variation within the object and background. It should be noted that this method is computationally intensive and convergence is not guaranteed (may converge to the boundary points of the grey level range).

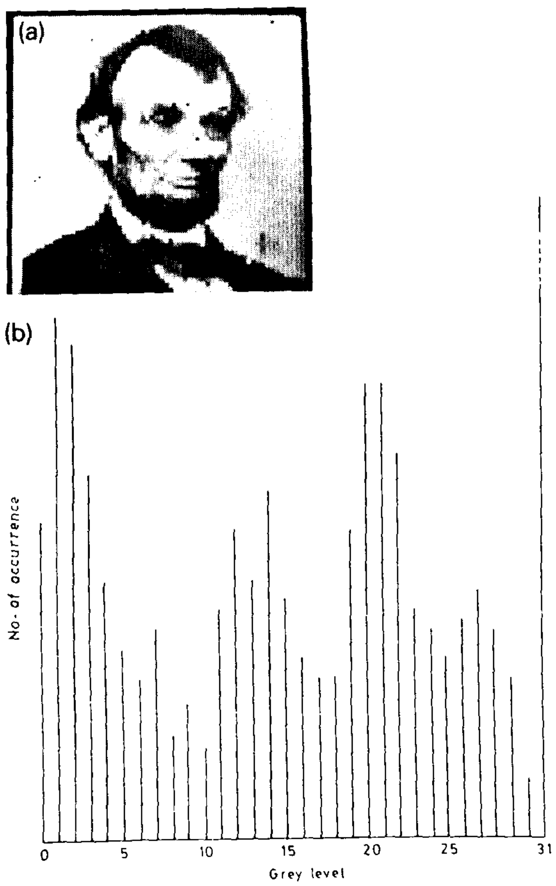
Pun (1980) and Kapur et al. (1985) have used entropy of the histogram of an image as the criterion for object-background classification. In Pun (1980), the a posteriori entropy of the partitioned image defined as

$$H_s = -P_s \log P_s - (1-P_s)\log(1-P_s)$$

(s is the assumed threshold and $P_s = \sum_{x_i=0}^{s} h(x_i)/MN$) is maximized to obtain a threshold for segmentation.

Kapur et al. (1985) considered two probability distributions, one for the object and the other for the background. The entropy of the partitioned image is then maximized to obtain a threshold for segmentation. In other words, they maximized

$$H_s = -\sum_{x_i=0}^{s} \frac{p(x_i)}{P_s}\log\frac{p(x_i)}{P_s} - \sum_{x_i=s+1}^{L-1} \frac{p(x_i)}{1-P_s}\log\frac{p(x_i)}{1-P_s}.$$



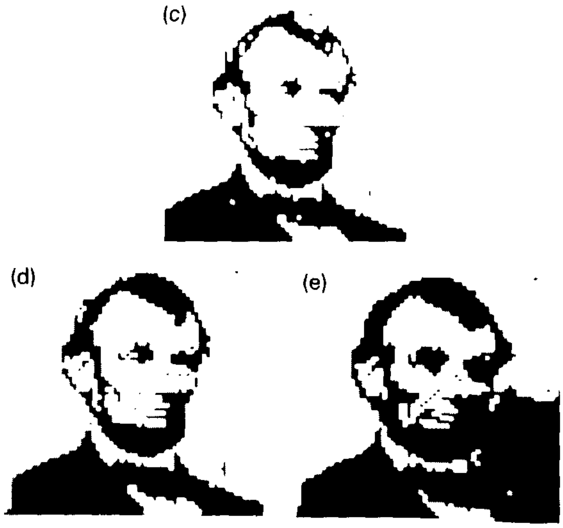


Figure 3. Lincoln image. (a) Input. (b) Histogram. (c) Output obtained using divergence. (d) Output obtained using dissimilarity assuming Poisson and non-parametric. (e) Output produced by the method of Kapur et al. (1985).

In the next part of this section we shall discuss the results obtained by the proposed and existing algorithms (Kittler and Illingworth (1986), Pun (1980), Kapur et al. (1985)).

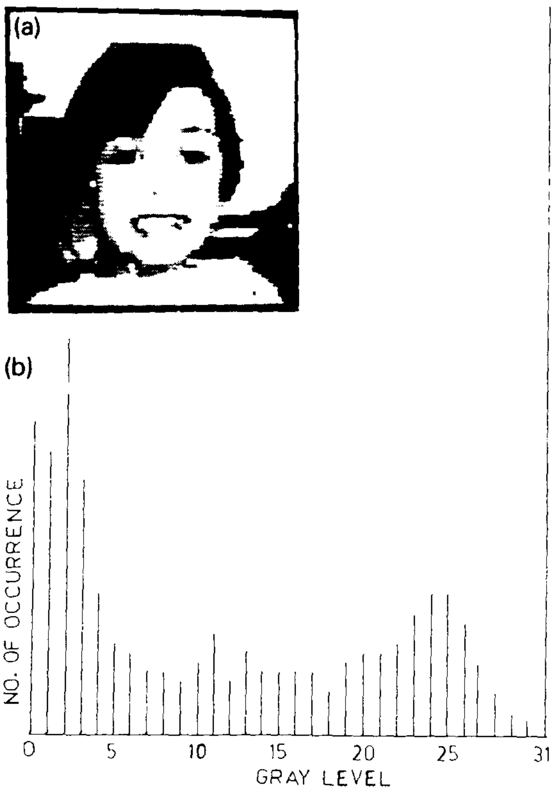
Table 1 shows the thresholds obtained by the suggested algorithms. Figures 2(a) and 2(b) represent the input image of a biplane and its histogram, respectively. From the table, one can observe that all the methods produce comparable thresholds. The outputs obtained using the divergence and the dissimilarity measure are shown in Figures 2(c)–(e).

Figure 3(a) represents the input image of Abraham Lincoln with a multimodal histogram (Figure 3(b)). For this type of image, multi-thresholding is more appropriate. But, since this image has two clear portions (object and background), an attempt has been made to find the best possible partitioning. Here, the divergence measure and the dissimilarity measure with Poisson parameters resulted in good segmentations (Figures 3(c)–(d)). For this image, the dissimilarity measure using the normal distribution is not able to extract the object.

For the Boy image (Figure 4(a)), the segmented outputs obtained by different methods are shown in Figures 4(c)–(e). In this case the dissimilarity measure with a normal distribution fails completely.

To establish the effectiveness of the proposed methods we have implemented them on three test data which are shown in Figures 5(a)–(c). It is to be noted here that for these histograms all the methods produce good thresholds (see Table 1).

So as to have a comparative study, some of the existing thresholding techniques (Kittler and Illingworth (1986), Pun (1980), Kapur et al. (1985)) have also been implemented. The thresholds obtained by the algorithms are depicted in Table 2. The results produced by the algorithm of Kittler and Illingworth (1986) are not satisfactory except for the Biplane image, where it was able to segment properly (Figure 2(f)). It is also to be noted here that the algorithm does not converge for the



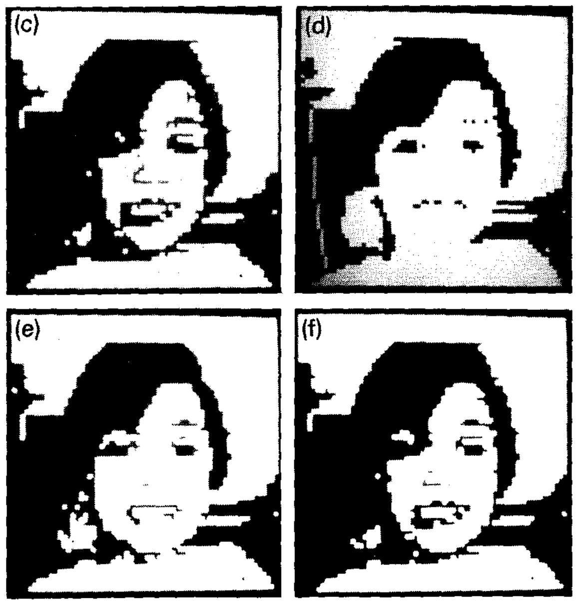


Figure 4. Boy image. (a) Input. (b) Histogram. (c) Output obtained using divergence. (d) Output obtained using dissimilarity assuming Poisson. (e) Output obtained using dissimilarity assuming non-parametric and method of Kapur et al. (1985). (f) Output produced by the method of Pun (1980).

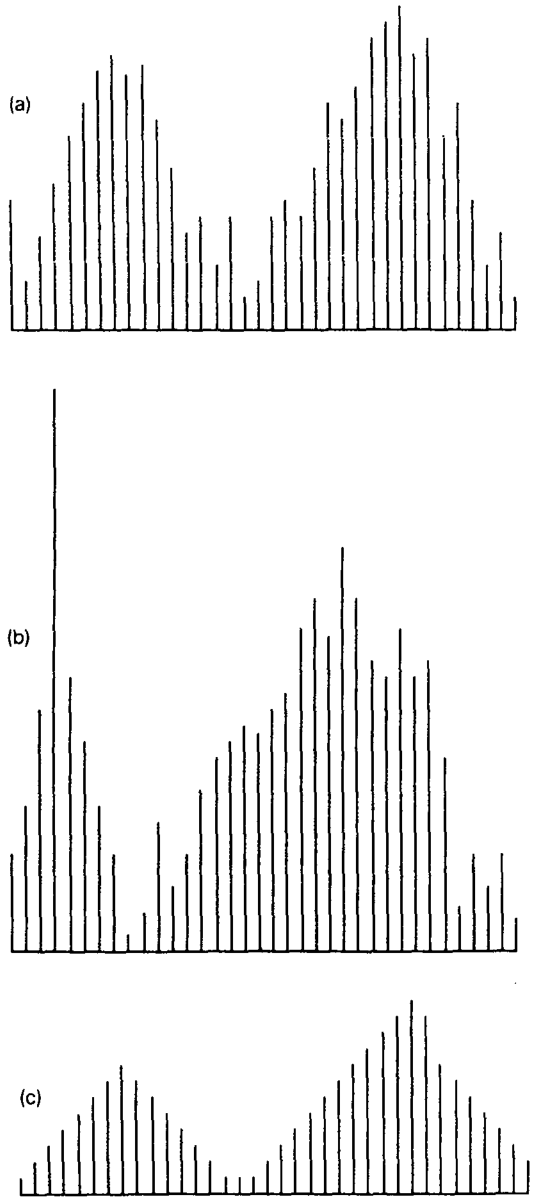


Figure 5. Histograms for (a) Test 1, (b) Test 2, (c) Test 3.

Lincoln and Boy images. The methods of Pun (1980) and Kapur et al. (1985) are also not able to extract the objects. Only the method proposed by Kapur et al. (1985) has produced a reasonable result (Figure 4(d)) for the Boy image. A visual inspection of the thresholded images shows that the thresholds obtained by the proposed algorithms are better (which can further be verified from the valleys of the histograms of the images).

Table 2

Images	Thresholds			
	Kittler & Illingworth		Method of	
	Initial	Final	Pun	Kapur
Biplane	5	2		
	5	10		
	10	10	25	22
	15	10		
	28	10		
Lincoln	2	2		
	5			
	10	-	17	16
	15	-		
	28	28		
Boy	2	2		
	5	-		
	10	-	16	14
	15	-		
	28	32		

Conclusions

A divergence measure between two fuzzy sets has been suggested, which satisfies all properties of a metric except the triangle inequality property. Some properties of this pseudo-metric have been discussed. This measure has been used to partition an image into object and background. Like divergence in probability theory this fuzzy divergence quantifies the discrepancy between two fuzzy sets.

A tailored version of the probability measure of a fuzzy event has been viewed as an index of similarity/dissimilarity between two sets and used to develop algorithms for image segmentation. In this context the grey level histogram of the image has been considered as a mixture of two probability distributions (may be parametric or non-parametric).

The algorithms have been applied to a set of three images and to some test data. Both measures, fuzzy divergence and dissimilarity, produced satisfactory results. Results have also been compared with three existing algorithms. It may be mentioned that the performance of the proposed methods is better for images with bimodal histograms.

References

Kandel, A. (1982). *Fuzzy Mathematical Techniques with Applications*. Addison-Wesley, Reading, MA.

Kapur, J.N., P.K. Sahoo and A.K.C. Wang (1985). A new method for grey level picture thresholding using entropy of the histogram. *Computer Graphics, Vision, and Image Processing* 29, 273–285.

Kittler, J. and J. Illingworth (1986). Minimum error thresholding. *Pattern Recognition* 19(1), 41–47.

Kullback, S. (1959). *Information Theory and Statistics*. Wiley, New York.

Murthy, C.A. and S.K. Pal (1990). Fuzzy thresholding: mathematical framework, bound functions and weighted moving average technique. *Pattern Recognition Lett.* 11, 197–206.

Pal, N.R. and D. Bhandari (1992). Object-background classification: some new techniques. *Signal Processing*, to appear.

Pal, S.K. and D. Dutta Majumder (1986). *Fuzzy Mathematical Approach to Pattern Recognition*. Wiley, New York.

Pal, N.R. and S.K. Pal (1991). Image model, Poisson distribution and object extraction. *Int. J. Pattern Recognition and Artificial Intelligence* 5(3), 459–483.

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

Zadeh, L. (1986). Probability measures of fuzzy events. *J. Math. Anal. Appl.* 23, 421–427.