

Possibility Expectation and Its Decision Making Algorithm**James M. Keller and Bolin Yan**Electrical and Computer Engineering
University of Missouri
Columbia, MO 65211, USA

52-61

143 401

p-16

Abstract

The fuzzy integral has been shown to be an effective tool for the aggregation of evidence in decision making. Of primary importance in the development of a fuzzy integral pattern recognition algorithm is the choice (construction) of the measure which embodies the importance of subsets of sources of evidence. Sugeno fuzzy measures have received the most attention due to the recursive nature of the fabrication of the measure on nested sequences of subsets. Possibility measures exhibit an even simpler generation capability, but usually require that one of the sources of information possess complete credibility. In real applications, such normalization may not be possible, or even desirable. In this report both the theory and a decision making algorithm for a variation of the fuzzy integral are presented. This integral is based on a possibility measure where it is not required that the measure of the universe be unity. A training algorithm for the possibility densities in a pattern recognition application is also presented with the results demonstrated on the shuttle-earth-space training and testing images.

1. Introduction

Decision making is a basic problem in science, engineering, and even in daily life. There are often conflicting requirements of low error rates and minimum computation time to reduce the cost. The purpose of this paper is to propose the concept of possibility expectation via the possibility integral as a decision making scheme, which can be used to construct optimal decision making algorithms. A possibility expectation is a value of nonlinear integration of two pieces of information, namely, an evidence function $h(x)$ and a possibility measure $\text{Pos}(\cdot)$. A possibility measure is a monotonic set function with the property that the measure of the universe X can be less than or equal to unity.

An example of possibility expectation is the following: In the court room, although the witnesses for both the defendant and plaintiff promise that they will tell the truth, the judge still needs to assign the grade of credibility (possibility densities) to each person to evaluate what the person says (evidence). The judge will integrate what each group of witnesses said with his belief in that group's credibility (possibility measure). Then the judge makes his decision

(possibility expectation).

In multicriteria decision making, as can be found in most pattern recognition problems, the value of each source of information (and thus all subsets of sources) toward each alternative can be different. For example, "greenness" may be a very important feature for recognizing certain types of trees in an image; whereas it may be quite unimportant as a feature for a roof of a building. This difference in the importance or credibility of subsets of information sources will be encoded in a possibility measure. The degree to which a given image region is green, to continue the example, is objective evidence supplied by the information source. After collecting all such objective information, it is the job of the decision making algorithm to fuse the objective evidence together with the worth of the sources. In our methodology, this will be accomplished by utilizing the possibility integral, a variation of the fuzzy integral [1].

The particular possibility measures which we describe generalize fuzzy measures in that it is not required that the measure of the entire domain of discourse be one. In a pattern recognition problem, it may not be possible, or may not be desirable to force one of the sources of information to have "perfect credibility". By relaxing this requirement, not only do we match real situations better, we also provide the opportunity to create better decision making algorithms, as we shall see later.

For a pattern recognition environment, a method to learn the possibility densities (values upon which the measure is generated) from training data is given. The results of the subsequent algorithm are used to segment a shuttle from the earth and space background.

2. Possibility Measures and Possibility Integral

Definition 2.1 A set function $\text{Pos}(\cdot): 2^X \rightarrow [0, 1]$ is called a possibility measure if it satisfies the following properties:

- (1) $\text{Pos}(\emptyset) = 0, \text{Pos}(X) \leq 1.$
- (2) If $A, B \in 2^X$ and $A \subset B$, then $\text{Pos}(A) \leq \text{Pos}(B).$
- (3) $\text{Pos}(\bigcup_{j=1}^n A_j) = \sup_{\mu \in [1, n]} [\text{Pos}(A_j)].$

Note: If X is finite, a possibility measure is not a fuzzy measure when $\text{Pos}(X) < 1$; it is the same as fuzzy measure only when $\text{Pos}(X) = 1$. If X is infinite, a possibility measure is not a fuzzy measure in general [2]. Puri and Ralescu [3] give two counterexamples which show that, even in "nice" cases, a possibility measure is not a fuzzy measure in the infinite case.

Definition 2.2 Let $X = \{x_j \mid j = 1, \dots, n\}$ be a finite set and let Pos be a possibility measure on 2^X . The set $\{p^j = \text{Pos}(\{x_j\}) \mid j = 1, \dots, n\}$ is called the set of possibility densities for Pos .

By definition of the possibility measure, it is clear that the measure of any subset A of X can be generated by

$$\text{Pos}(A) = \max_{x_j \in A} \{p^j\},$$

and hence, a possibility measure is easily generated by its densities.

We note that possibility theory can be induced not only from the nested bodies of evidence within the Dempster-Shafer theory [4], but also from the fuzzy sets introduced by Zadeh [6]. A fuzzy set F is a set whose elements are characterized by the membership grade function $\mu_F(x): X \rightarrow [0, 1]$. A value of $\mu_F(x)$ expresses the grade of membership that an element $x \in X$ belongs to the fuzzy subset F of X . Let $\pi_F(x) = \mu_F(x)$ be a possibility distribution induced by a fuzzy set F . In general, a possibility distribution is thought of as an elastic restriction on the values within a domain of discourse which a fuzzy variable may assume [5]. The fuzzy set F provides the meaning of the restriction. A possibility measure is defined as $\text{Pos}(A) = \sup_{x \in A} [\pi_F(x)]$ for all $A \in 2^X$. This relationship holds also for non-normal fuzzy sets [6]. Although a fuzzy set and a possibility distribution have a common mathematical expression, the underlying concepts are different [5].

Our possibility measures are non-normalized generalizations of what are referred to as S-decomposable measures [7, 8], these being a class of fuzzy measures which are easily computable.

Definition 2.3 Let $h(x)$ be a function such that $h: X \rightarrow [0, 1]$, and let $\text{Pos}(\cdot)$ be a possibility measure of 2^X . The possibility integral or the possibility expectation of $h(x)$ with respect to $\text{Pos}(\cdot)$ is defined as

$$\int_X h(x) \circ \text{Pos}(\cdot) = \sup_{\alpha \in [0, 1]} [\alpha \wedge \text{Pos}(A_\alpha)], \quad \text{where } A_\alpha = \{x \mid h(x) \geq \alpha\}.$$

When $X = \{x_i \mid i = 1, \dots, n\}$ is finite, if we reorder X such that $h(x_1) \geq h(x_2) \geq \dots \geq h(x_n)$, then the possibility integral can be written as

$$\int_X h(x) \circ \text{Pos}(\cdot) = \bigvee_{j=1}^n [h(x_j) \wedge \text{Pos}(A_j)], \quad \text{where } A_j = \{x_1, x_2, \dots, x_j\}.$$

The rationale of the possibility expectation is to find the source within the universe where both the information value $h(x_j)$ and the possibility measure $\text{Pos}(A_j)$ are compatibly large, that is, where the feasibility of the data and the reliability of a subset of sources is jointly optimal.

The fuzzy integral developed by Sugeno [1] has the same formulation with the exception that a fuzzy measure is used in lieu of the possibility measure. One of the advantages of the possibility integral is that the measures $\text{Pos}(A_j)$ are easily calculated from the densities by the recursive relationship

$$\text{Pos}(A_1) = \text{Pos}(\{x_1\}) = p^1;$$

$$\text{Pos}(A_j) = \text{Pos}(A_{j-1} \cup \{x_j\}) = \text{Pos}(A_{j-1}) \vee p^j.$$

In contrast, for Sugeno fuzzy measure g_λ with the fuzzy densities $\{g^1, \dots, g^n\}$, this recursive definition becomes

$$g_\lambda(A_1) = g_\lambda(\{x_1\}) = g^1;$$

$$g_\lambda(A_j) = g_\lambda(A_{j-1} \cup \{x_j\}) = g_\lambda(A_{j-1}) + g^j + \lambda g^j g_\lambda(A_{j-1}),$$

where $\lambda \geq -1$ [1, 10, 11]. The value of λ must be calculated from the equation

$$\prod_{i=1}^n (1 + \lambda g^i) = 1 + \lambda, [1].$$

If one is going to try to learn a measure (iteratively) from training data, the amount of computations necessary to learn a possibility measure, and then evaluate its possibility integral is considerably less than that required for a Sugeno fuzzy measure and its fuzzy integral.

For a multiclass pattern recognition problem (or any multicriteria decision making problem), the set \mathbf{X} represents sources of information (criteria). Each class (alternative) will have its own evidence function $h_i: \mathbf{X} \rightarrow [0, 1]$ to assess the feasibility that the decision is class i (alternative i) from the standpoint of each individual source, x_j . Also, each class will have its own possibility measure Pos_i which determines the worth of all subsets of sources in deciding that a particular object belongs to class i . Finally, the collection of possibility integrals

$$e_i = \int_{\mathbf{X}} h_i(x) \circ \text{Pos}_i(\cdot),$$

gives a class-individualized "fusion" of the direct evidence with the worth of that evidence. A final crisp decision can be made from the possibility expectations (integral values), for example, pick the class corresponding to the maximum possibility expectation. Alternately, these expectation values can be used as confidences for later processing.

3. Properties of The Possibility Integral

Several interesting properties of the possibility integral are proved in [11]. Of particular

interest to the algorithm presented in the next section are the following two results.

Theorem 3.1 $0 \leq \int_X h(x) \circ \text{Pos}(\cdot) \leq \text{Pos}(X).$

Theorem 3.2 If $h_1(x) \leq h_2(x) \forall x$:

$$\begin{aligned} \int_X h_1(x) \circ \text{Pos}(\cdot) &\leq \int_X h_2(x) \circ \text{Pos}(\cdot), & \text{if } \text{Pos}(X) > h_1(x) \text{ for all } x, \\ \int_X h_1(x) \circ \text{Pos}(\cdot) &= \int_X h_2(x) \circ \text{Pos}(\cdot), & \text{if } \text{Pos}(X) \leq h_1(x) \text{ for all } x. \end{aligned}$$

4. Decision Rule and Training Algorithm

In the procedure given below, we consider a two class pattern recognition problem, or a two alternative decision process. The approach can be extended directly to multiple classes, but from the particular structure of the training mechanism, it would be more appropriate to view it as a series of two class problems, either as pairwise distinctions, or as each class against all of the remaining classes. Since the possibility integral algorithm does not create geometric decision boundaries in feature spaces (as, for example, Bayes Decision Theory), the second approach is reasonable and contains fewer subdecisions which need to be made to extend this to multiple classes.

The actual decision algorithm utilizes the nature of the possibility integral to split the input objects (as represented by the evidence function $h(x)$) into four groups to reduce the computational load. The first two groups deal with the case where the strength of all objective evidence for one class outweighs that for the other. In most cases, this corresponds to the fact that, in a pattern recognition problem, a majority of the data are easily distinguished (being quite typical of their class). Decision rules 1 and 2 below are a consequent of Theorem 3.2 assuming that the possibility measures for both classes in this case are identical. Of course, there are problems where the objective evidence for one class can dominate that for the other class, and yet, the object belongs to the latter. This could happen if the worth of the source, i.e., the densities, are vastly different between classes. During training, this condition is monitored, and if the training data produce such outcomes, the first two rules are abandoned, forcing all training samples to be "conflict data".

The initial definition of "conflict" is an object where the evidence function for one class does not dominate that of the other. In this case, we split the training data (and also the unknown test objects) into two subgroups based on the class receiving the highest degree of support from any source. For each group, two possibility measures are formed which minimize the total misclassification of the training data. The purpose of partitioning the data in this

manner is to reduce the size of the training set since our initial training scheme is a complete search through a quantized set of all pairs of density functions. To reduce further the amount of computations, we note that the value of a possibility integral cannot be larger than the maximum of the function being integrated. This fact allows us to restrict the range of density values to be no larger than the maximum evidential support in the training set. (Reducing the training sets gives more opportunity to invoke this restriction). Optimal pairs of density functions (in term of minimal error rate on the training data) are formed and then used in the testing cycle. There are 4 possibility measures generated during training - one from each class in each of the two subgroups of conflict data.

The decision algorithm is summarized below.

BEGIN

FOR each feature data vector DO obtain $h_1(x_j)$ for all j and $h_2(x_j)$ for all j ;

(1) IF $h_1(x_j) > h_2(x_j)$ for all j , THEN the feature data vector belongs to class 1.

(2) ELSE IF $h_1(x_j) < h_2(x_j)$ for all j , THEN the feature data vector belongs to class 2.

(3) ELSE

 If $\bigvee_{j=1}^n h_1(x_j) \geq \bigvee_{j=1}^n h_2(x_j)$, Then

$$e_1 = \bigvee_{j=1}^n [h_1(x_j) \wedge \text{Pos}_{11}(A_j)], \quad e_2 = \bigvee_{j=1}^n [h_2(x_j) \wedge \text{Pos}_{12}(A_j)]$$

 Else

$$e_1 = \bigvee_{j=1}^n [h_1(x_j) \wedge \text{Pos}_{21}(A_j)], \quad e_2 = \bigvee_{j=1}^n [h_2(x_j) \wedge \text{Pos}_{22}(A_j)]$$

 End If

 If $e_1 > e_2$, Then the feature data vector belongs to class 1,

 Else the the feature data vector belongs to class 2.

 End If

END IF

END FOR

END .

5. Experimental Results

Two shuttle-earth-space intensity images were used in the experiment, in which all the data from the two images were treated as "conflict data" and hence only the third decision rule applies.

The training image is shown in Fig. 5.4 and the test image is shown in Fig. 5.5. Three texture feature images (contrast, difference, and the entropy) were derived from the training

and the test images respectively, i.e., three feature images for training and three feature images for testing (For the definition of these features, please see section on membership generation techniques in this report). The three feature images, used for training the possibility densities, are shown in Fig. 5.2. The three feature images used in testing are shown in Fig. 5.6.

The possibility distribution (or membership function) of each class in each feature, that used to generate the evidential function $h(x)$, is determined by using the possibilistic clustering algorithm on the histograms of each class in each feature, which is described in another section of this report.

While training, the possibility densities were determined with the "perceptron criterion" (i.e., minimize the decision error) from the feature images in Fig. 5.2. The segmentation result corresponding to the possibility measure(s) for the training image is shown in Fig. 5.3, in which the shuttle and its background are clearly segmented, except that the shuttle body seems disconnected. To improve the connection of the shuttle body, the possibility densities of the shuttle were raised slightly, from which the segmentation result in Fig. 5.4 and the result in Fig. 5.7 (for the test case) were obtained. These results can be improved quite easily with a shrink-expand operation.

6. Conclusion

In this paper, a decision making algorithm based on a variation of the fuzzy integral was proposed. The possibility integral has a particularly simple generation capability. The algorithm was run on the shuttle-earth-space images, reasonable good results were obtained.

7. References

- [1] M. Sugeno, "Theory of Fuzzy Integrals and its Applications", *Ph.D. Thesis*, Tokyo Institute of Technology, 1974.
- [2] A. Kandel, *Fuzzy Mathematical Techniques with Applications*, Addison-Wesley: Reading, MA, 1986.
- [3] M. Puri and D. Ralescu, "A Possibility Measure is Not a Fuzzy Measure", *Fuzzy Sets and Systems*, Vol. 9, No. 3, pp. 219-227.
- [4] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press: Princeton, NJ, 1976.
- [5] L. A. Zadeh, "Fuzzy Sets as a Basis for a Theory of Possibility", *Fuzzy Sets and Systems*, Vol. 1, No. 1, 1978, pp. 3-28.
- [6] D. Dubois and H. Prade, *Possibility Theory*, Plenum Press, 1986, pp. 9.
- [7] S. Weber, "Two Integrals and Some Modified Versions- Critical Remarks", *Fuzzy Sets and Systems*, Vol. 20, 1986, pp 97-105.
- [8] H. Tahani, "Generalized Fuzzy Integral in computer vision", Ph.D. Thesis, University of Missouri-Columbia, 1991.
- [9] B. Yan and J. Keller, "Conditional Fuzzy Measure and Image Segmentation", *Proceedings of North American Fuzzy Information Processing Society* Columbia, Missouri, USA, 1991, pp. 32-36.
- [10] B. Yan, "Fuzzy Integral at $\lambda = -1$ ", *Proceedings of North American Fuzzy Information Processing Society*, Columbia,

Missouri, USA, 1991, pp. 18- 22.

- [11] B. Yan and J. Keller, "Possibility Integral". *Academic Research Report* University of Missouri-Columbia, 1991.

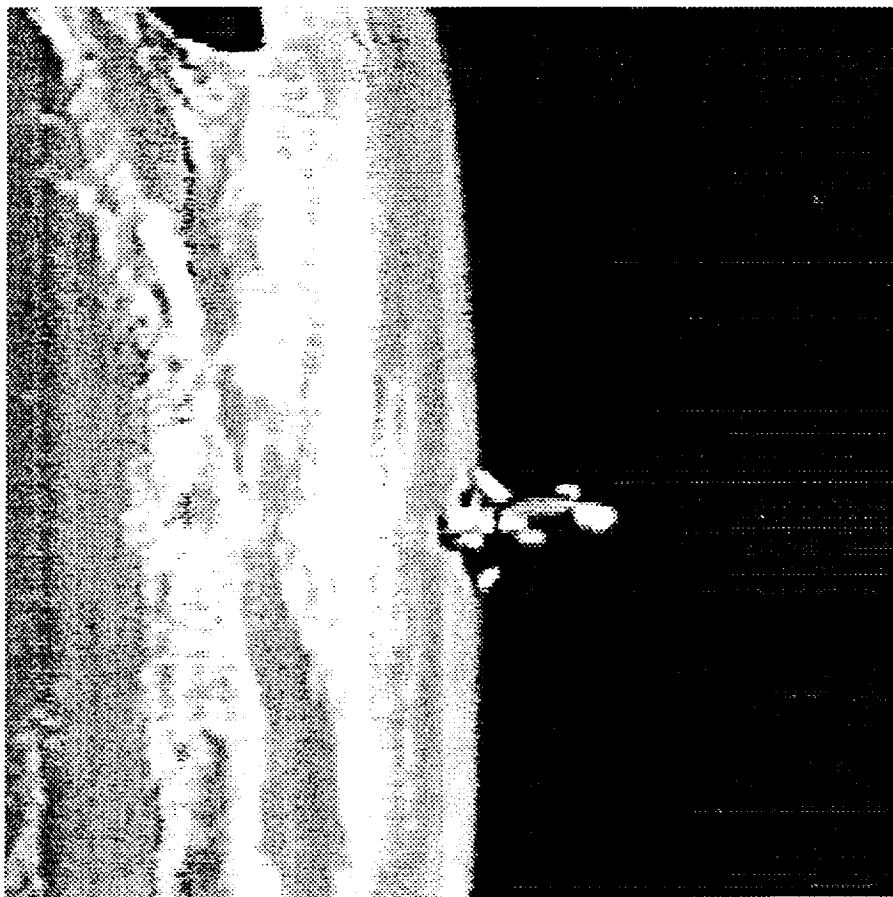


Fig 5.1 Intensity training image.

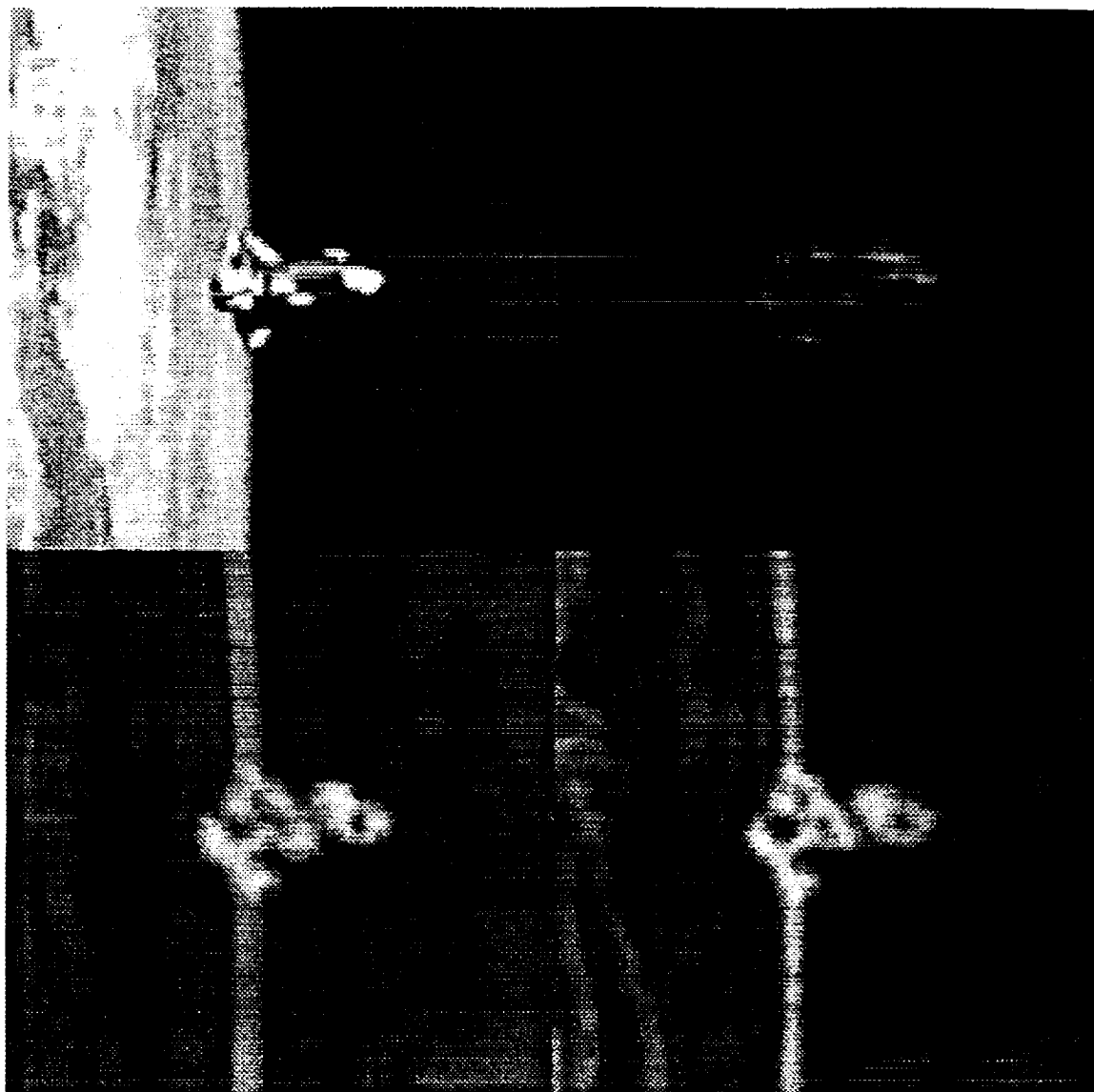


Fig 5.2 (top left) Intensity training image.
(top right) Contrast feature image.
(bottom left) Difference feature image.
(bottom right) Entropy feature image.

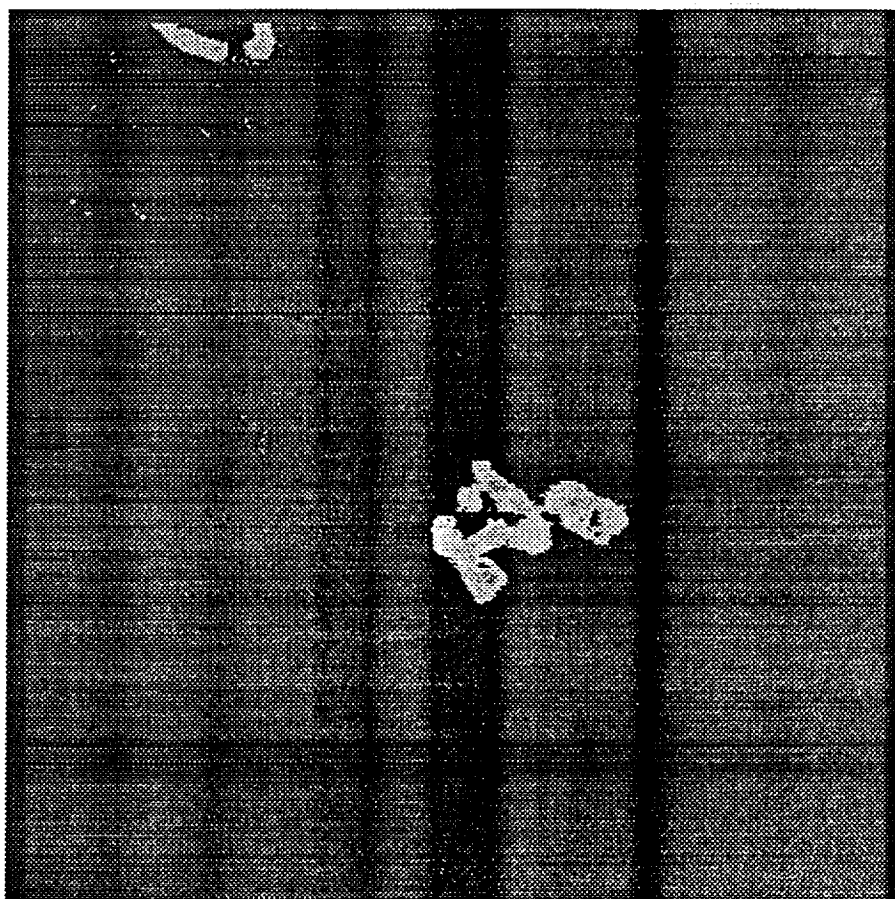


Fig 5.3 Segmented image1 using the possibility integral algorithm.

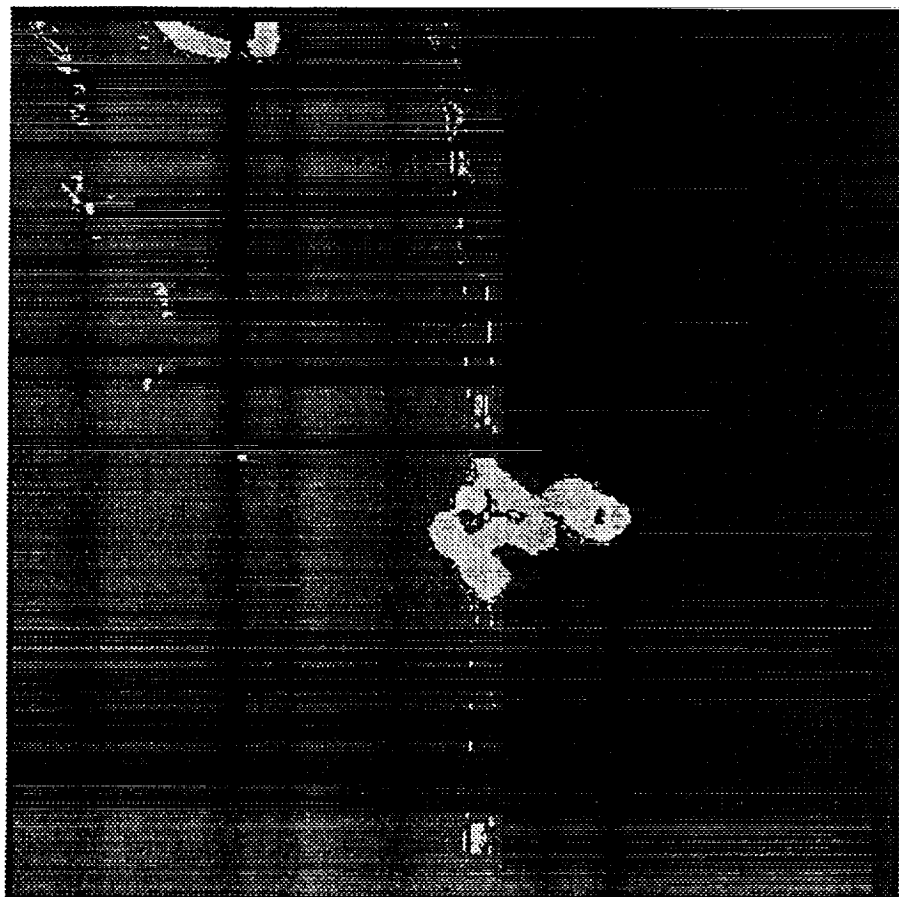


Fig 5.4 Segmented image2 using the possibility integral algorithm.

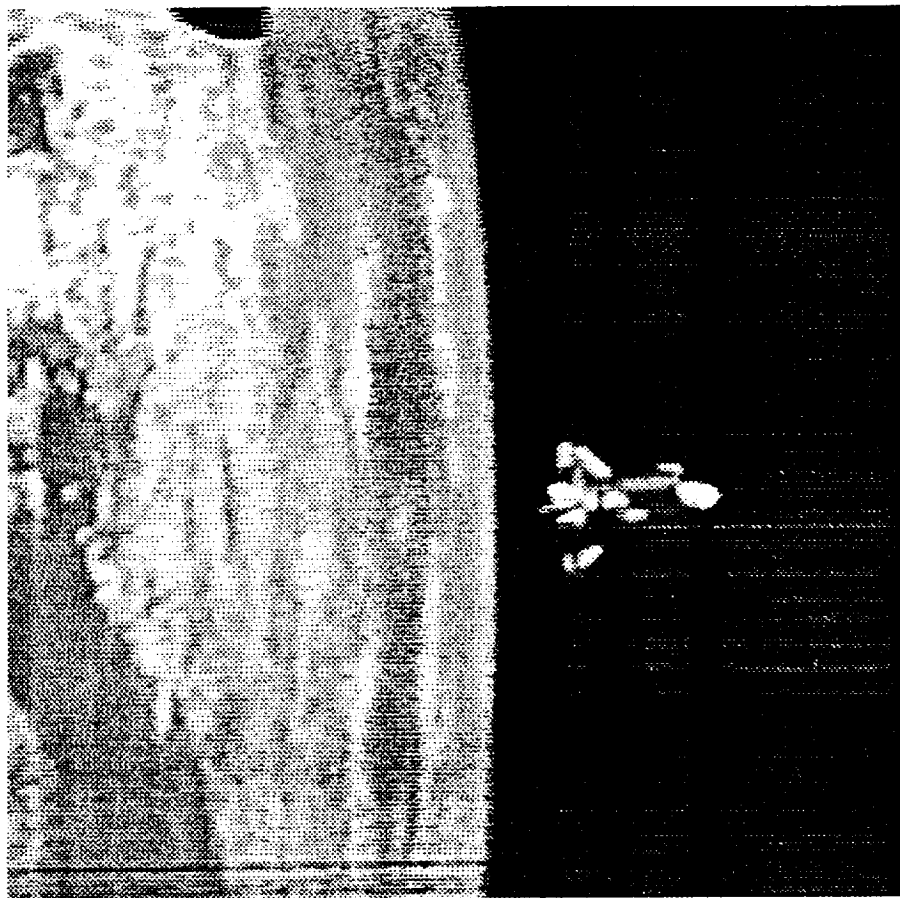


Fig 5.5 Intensity testing image.

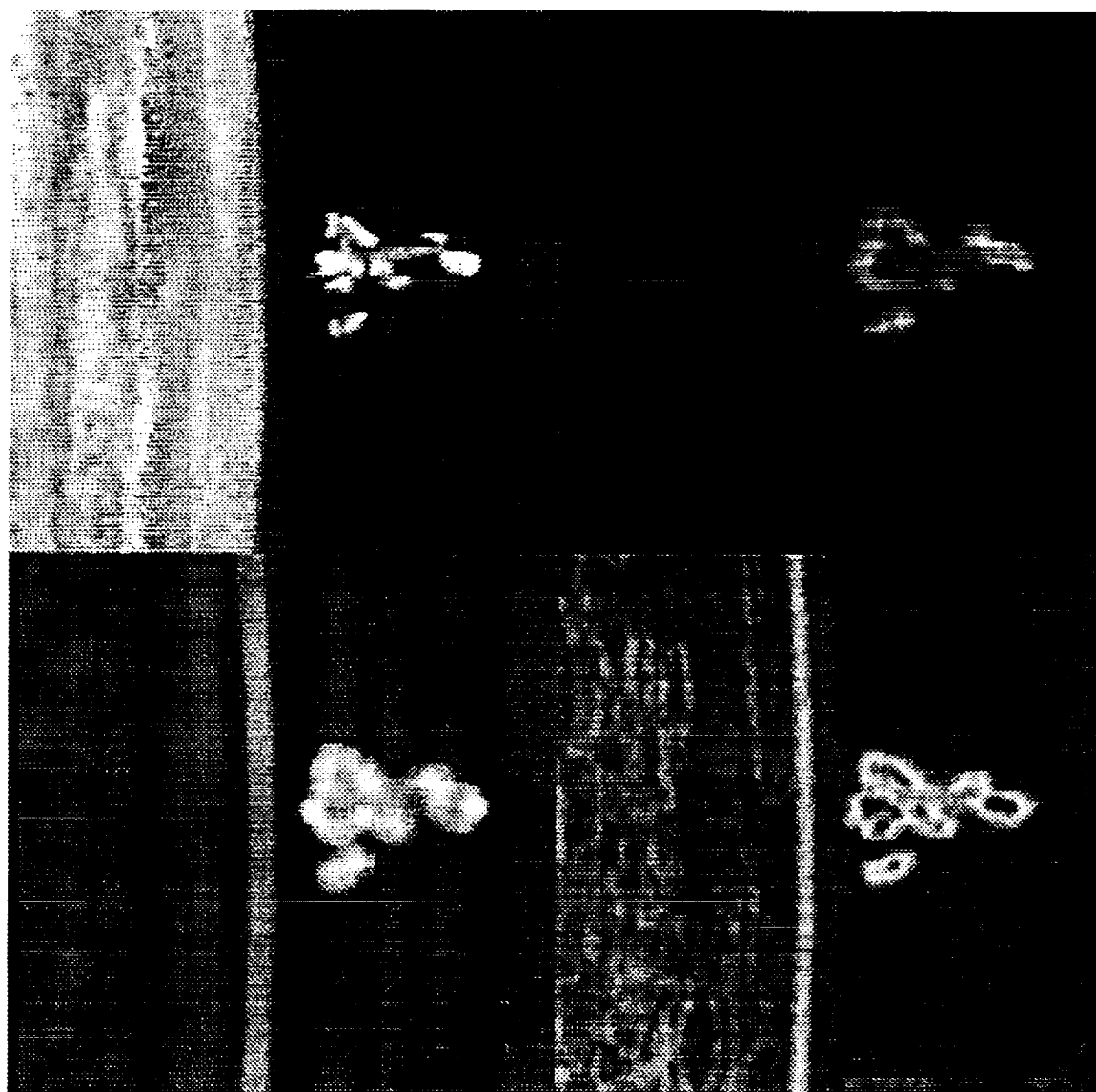


Fig 5.6 (top left) Intensity testing image.
(top right) Contrast feature image.
(bottom left) Difference feature image.
(bottom right) Entropy feature image.

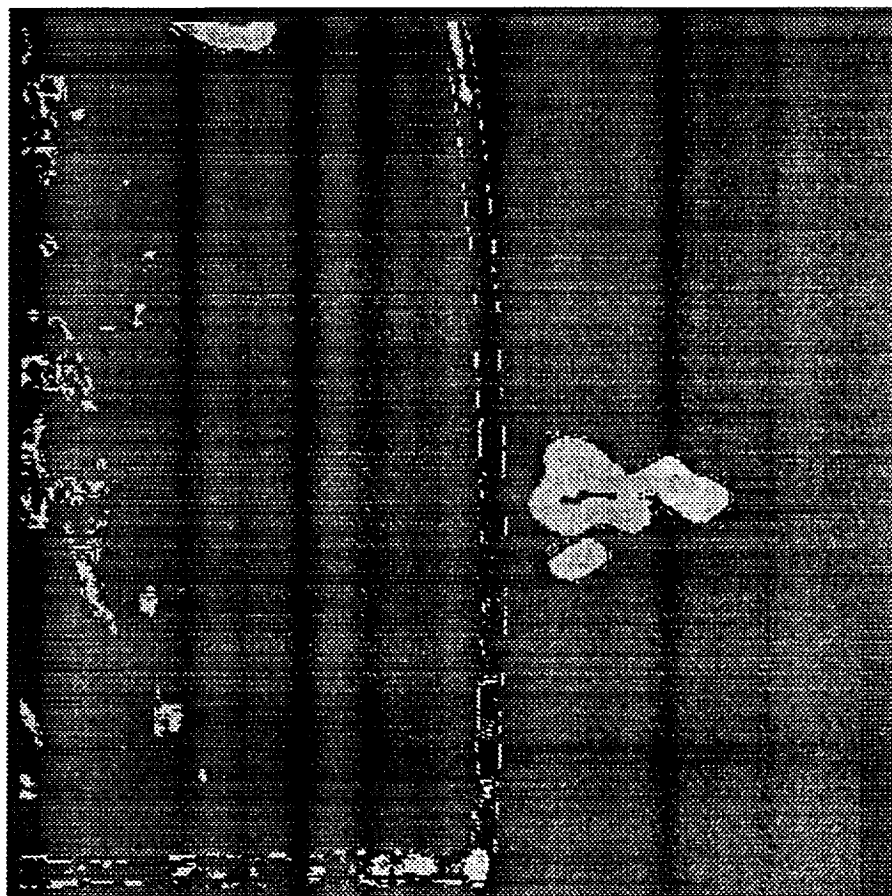


Fig 5.7 Segmented testing image using the possibility integral algorithm.

Calculation of Membership Functions

Our work in this area has progressed nicely. We have designed and implemented a new algorithm to generate membership values from a set of training data using a multi-layer neural network. This is in addition to the progress we made in the transformation of "probability density functions" into possibility distributions for use in assigning membership values to individual points as reported in the third quarter report.