ELSEVIER

Fuzzy Sets and Systems 65 (1994) 273–283

FUZZY sets and systems

Advances in fuzzy integration for pattern recognition

James M. Keller*, Paul Gader, Hossein Tahani, Jung-Hsien Chiang, Magdi Mohamed

Department of Electrical and Computer Engineering, University of Missouri-Columbia, Columbia, MO 65211, USA

Received February 1994; revised March 1994

Abstract

Uncertainty abounds in pattern recognition problems. Therefore, management of uncertainty is an important problem in the development of automated systems for the detection, recognition, and interpretation of objects from their feature measurements. Fuzzy set theory offers numerous methodologies for the modeling and management of uncertainty. One such fuzzy set theoretic technology which has proven quite useful in pattern recognition is the fuzzy integral. The purpose of this paper is to examine new utilizations of the fuzzy integral as a decision making model in the area of object recognition. In particular, we develop generalizations of the fuzzy integral and show that these generalizations can achieve higher recognition rates in an automatic target recognition problem. Also, we demonstrate significant increases in recognition rates using the fuzzy integral to fuse the results of different neural network classifiers in a complex handwritten character recognition domain.

Key words: Multiple criteria evaluation; Operators; Image processing; Engineering applications; Fuzzy integral; Densities; Pattern recognition

1. Introduction

Fuzzy set methods have recently achieved a high degree of success and popularity in many areas such as control, multi-criteria decision making, approximate reasoning, and pattern recognition [2, 3, 21, 36, 39]. What makes fuzzy set theory and fuzzy logic so attractive is the fact that they provide a powerful and flexible framework to represent vague and ill-defined concepts, and they represent and numerically manipulate linguistic rules in a natural way. Fuzzy methods are not a solution to all problems. However, they are useful in situations in which the concepts (features, criteria or rules) are vague. This is often the situation in pattern recognition.

The fuzzy integral is a numeric-based approach which has been used for both pattern recognition and image segmentation [12–14, 16–19, 25, 26, 30, 37, 38]. It uses a hierarchical network of evidence sources to arrive at a confidence value for a particular hypothesis or decision. A distinguishing characteristic of the fuzzy integral is that it utilizes information concerning the worth or importance of the sources in the decision making process. In following sections, we describe the basic formulation of the fuzzy integral and its generalizations. We then describe several experiments performed on nontrivial data sets in which the fuzzy integral

* Corresponding author.

SSDI 0165-0114(94)00102-D

increases pattern recognition performance. In an automatic target recognition (ATR) application, we first report earlier very good results using the standard formulation of the fuzzy integral and then apply the generalized fuzzy integral to the same data, showing how the classification performance can be further enhanced. Finally, we consider the issue of fusion of independent classifier outputs by the fuzzy integral. The ATR problem is revisited from a multisensor standpoint, and then a complex handwritten character recognition situation is examined where the fuzzy integral provides significant improvement in fusing the results from separate neural networks.

2. The fuzzy integral

In this section we briefly describe the basic fuzzy integral and give an example of its use in pattern recognition. We discuss the issues of generating the objective evidence as well as generating the densities for the fuzzy measures necessary to apply this technique.

The fuzzy integral is a nonlinear approach to combine multiple sources of uncertain information as is the case in automated pattern recognition. In these applications, the integral is evaluated over a set of information sources (sensors, algorithms, features, etc.) and the function being integrated supplies a confidence value for a particular hypothesis or class from the standpoint of each individual source of information.

The fuzzy integral relies on the concept of a fuzzy measure [4, 28, 32] which generalizes the concept of a probability measure. A fuzzy measure over a set X is a function:

$$g: 2^X \to [0, 1]$$

such that

(1) $g(\emptyset) = 0$; $g(X) = 1$;

(2) $g(B) \geqslant g(A)$ if $B \supset A$;

(3) If $\cdots \supset A_n \supset \cdots \supset A_1$, then $\lim g(A_i) = g(\bigcup A_i)$ (continuity).

A particularly useful class of fuzzy measures is due to Sugeno [28]. A fuzzy measure g_λ is called a Sugeno measure if it satisfies the following additional property for some $\lambda > -1$:

If $A \cap B = \emptyset$,

then $g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B)$.

If $\lambda = 0$, then g_λ is a probability measure. Suppose X is a finite set of information sources, $X = \{x_1, \ldots, x_n\}$, and let $g^i = g_\lambda(\{x_i\})$. The values $g^1, g^2, \ldots, g^n$, are called the fuzzy densities. The densities are interpreted as the importance of the individual information sources. The measure of a set A of information sources is interpreted as the importance of that subset of sources toward answering a particular question (such as class membership).

Using the above definitions one can show that $g_\lambda(A)$ can be constructed from the fuzzy densities of the elements of A for any subset A of X. Given the set of densities, the value of λ can be easily found as the unique root greater than -1 of a simple polynomial [28]. Thus, estimating the densities is a core problem when using the Sugeno measures, and, as will be seen later, many other classes of measures.

Let $h: X \to [0, 1]$. For example, X could be a set of individual feature types or even simple classifiers and $h(x)$ the confidence provided by feature (or classifier) x that an input sample is from a particular class. The Sugeno fuzzy integral of h over X with respect to g is defined in [28] by

$$e = \int h(x) \circ g = \sup_{\alpha \in [0, 1]} \{\alpha \wedge g(F_\alpha)\}$$

where $F_\alpha = \{x \mid h(x) \geqslant \alpha\}$.

In general, the set X is the set of information sources (algorithms, features, etc.) and the function h supplies a confidence value for a particular hypothesis or class from the standpoint of each individual source of information. The fuzzy measure supplies the expected worth of each subset of sources for this hypothesis.

In applications to pattern recognition, the computational cost of computing the confidence value e can be reduced significantly since the set of information sources is finite. If $X = \{x_1, \ldots, x_n\}$ is arranged so that $h(x_1) \geqslant h(x_2) \geqslant \cdots \geqslant h(x_n)$, then

$$e = \bigvee_i [h(x_i) \wedge g(X_i)]$$

where $X_i = \{x_1, \ldots, x_i\}$. This reduces the number of subsets needed to evaluate the fuzzy integral for each function h from 2^n down to just n. Also, the values $g(X_i)$ can be determined recursively from the definitions [28]. Of course, the function h must be sorted first, adding some complexity to the evaluation. The reader is refered to [4, 12, 28, 32] for more extensive theoretical background on fuzzy measures and the fuzzy integral.

In comparison with probability theory, the fuzzy integral corresponds to the concept of expectation. The fuzzy integral values provide a different measure of certainty in the classification than posterior probabilities. Since the integral evaluation need not sum to one, lack of evidence and negative evidence can be distinguished. Dempster–Shafer belief theory [27] can also distinguish between lack of evidence and negative evidence. A conceptual difference between the Sugeno fuzzy integral and a Dempster–Shafer classifier is in the frame of discernment. For the fuzzy integral, the frame of discernment contains the knowledge sources related to the hypothesis under consideration, whereas with belief theory, the frame of discernment contains all of the possible hypotheses. Thus the fuzzy integral algorithm has a means to assess the importance of all groups of knowledge sources towards answering the questions as well as the degree to which each knowledge source supports the hypothesis.

The Sugeno fuzzy integral is more computationally efficient than a strict Dempster–Shafer approach. With belief theory, each knowledge source would have to generate a belief function over the power set of the set of hypotheses, which are then combined using Dempster's rule. This calculation can have exponential complexity with the number of hypotheses. With the fuzzy integral, the measure need only be calculated for n subsets (where n is the number of knowledge sources for each hypothesis). These measures are then combined with the objective evidence to produce the integral values. In many applications, the number of knowledge sources is considerably less than the number of hypotheses, or classes.

The behavior of the fuzzy integral in an application is heavily dependent on the densities, or more generally, on the individual fuzzy measures. Therefore estimation of the densities or the measures is very important. In some applications of the fuzzy integral, the densities can be supplied subjectively by an expert [16]. This subjective assignment approach may be the only method to assess the worth of non numeric sources of information, such as context or "intelligence" reports. In most pattern recognition problems, it is preferable to estimate the densities directly from training data. This can be done by estimating how well individual sources separate a given class from others on the training data [17, 26, 30, 37], or by using the fuzzy integral itself in the learning process in an optimization approach. Keller and Yan have used iterative search methods to find optimal sets of densities for possibility measures [18, 38], while Grabisch et al. [13, 14] looked at the problem of learning the entire measure. In [13], they showed that by using the Choquet integral [29], the optimization can be addressed by linear or quadratic programming methods. Using the Sugeno integral, simulated annealing or heuristic search methods were employed [14].

A pattern recognition problem is approached as follows using the fuzzy integral. Information sources are identified. These sources could be individual features, pattern classifiers, context information, etc. A fuzzy measure is generated for each pattern class using some measurement, either subjectively generated or estimated from training data, of the worth of all subsets of the information sources toward recognizing objects from that class. This generation of the measures is the training phase for the fuzzy integral approach. Given a pattern to be classified, an evidence function $h_i(x_j)$ is evaluated for each information source x_j and each class ω_i. The functions h_i are then integrated with respect to their corresponding class fuzzy measures, resulting in one confidence value for each class. These confidence values are used to make a final classification decision, e.g., assign the pattern to the class with the highest confidence.

As an example, consider a fuzzy integral-based classifier for automatic target recognition (ATR). A fuzzy integral classifier was developed and tested using forward looking infrared (FLIR) images containing two tanks and an armored personnel carrier (APC) [30]. There were three sequences of 100

Table 1(a)
Fuzzy integral classifier for a two class ATR problem

A. Computed densities and λ values					
	g^1	g^2	g^3	g^4	λ
Tank	0.16	0.23	0.19	0.22	0.760
APC	0.15	0.24	0.18	0.23	0.764

Table 1(b)

	Tank	APC
B. Results for the fuzzy integral classifier (Confusion matrix) (Total correct 92.6%)		
Tank	175	1
APC	17	49
C. Results for the Bayes classifier (Confusion matrix) (Total correct 90.9%)		
Tank	176	0
APC	22	44
D. Results for the Dempster–Shafer classifier (Confusion matrix) (Total correct 86.4%)		
Tank	160	16
APC	17	49

frames each used for training purposes. In each sequence, the vehicles appeared at a different aspect angle to the sensor (0°, 45°, 90°). In the fourth sequence the APC "circled" one of the tanks, moving in and out of a ravine and finally coming toward the sensor. This sequence was used to perform the comparison tests. The images were preprocessed to extract object of interest windows. The classification level integration was performed using four statistical features calculated from the windows, that is, the sources $\{x_1, x_2, x_3, x_4\}$ represent {"mean", "variance", "skewness", kurtosis"} of image neighborhoods. To get the partial evaluation, $h(x)$, for each feature, the fuzzy 2-means algorithm [1] was used. The fuzzy densities, the degree of importance of each feature, were assigned based on how well these features separated the two classes Tank and APC on training data [30]. The result of the fuzzy integral classifier is presented in the form of confusion matrix, in Table 1, where the count of samples listed in each row are those which belong to the corresponding class and the count of samples listed in each column are those after classification, which was performed by choosing the class with the largest integral value.

We compared results achieved using the fuzzy integral with those achieved using a Bayes and a Dempster–Shafer rule-based classifier [35]. These results are also shown in Table 1. The fuzzy integral achieved higher recognition rates than both.

In [30], we demonstrated, on the above data, the ability of the fuzzy integral to fuse the outputs of three classifiers: a Bayes recognizer, the fuzzy C-means, and a feature-level fuzzy integral. Here, the densities were chosen heuristically based on the individual classifier performance on a training set. It was shown that the integration process was able to "correct" mistakes made by one of the classifiers, while maintaining the correct classifications for those objects where there was no confusion in the algorithm outputs.

3. Generalized (Weber) fuzzy integrals

The basic notion of a fuzzy integral using the Sugeno measure has been demonstrated to be a useful tool. It can be improved by using more general measures or by using different fuzzy aggregation operators in the definition of the fuzzy integral. Weber [33] proposed the first generalization of the Sugeno integral. The most complete treatment of the theoretical aspects of these generalized fuzzy integrals can be found in [24]. Keller and Tahani have extended the fuzzy integral information fusion approach to a large family of measures, called S-decomposable measures [17, 31]. Given a triangular co-norm S, an S-decomposable measure g has the property [33]:

If $A \cap B = \emptyset$, then $g(A \cup B) = S(g(A), g(B))$.

Possibility measures [4] are simple examples of such S-decomposable measures where S is the maximum operator. Clearly, Sugeno measures are also examples of S-decomposable measures.

An important property of this class is that the measure of an arbitrary set of information sources can be computed if the densities are known, as with the Sugeno measures. Many other S-decomposable measures can actually be constructed by the definition from a set of density values for a given t-conorm S if the boundary conditions hold, i.e., one must guarantee that $g(X) = 1$. This will clearly happen if one of the densities has value 1 (a sufficient but not a necessary condition). This follows simply from the fact that

$$g(X) = g(\{x_1\} \cup \{x_2\} \cup \cdots \cup \{x_n\})$$
$$= S(g(\{x_1\}), g(\{x_2\}), \ldots, g(\{x_n\}))$$

and $S(1, y) = 1$ where $g(\{x_i\})$ is the density value for $i = 1, \ldots, n$.

In our applications, we normalized the densities to achieve this sufficient condition. Thus, estimating the measure is still reduced to an estimation of the densities. In [31], Tahani investigated the use of several t-conorms to generate S-decomposable measures, including the Schweizer and Sklar family of parameterized t-conorms given by

$$S_p(a, b) = (a^p + b^p - a^p b^p)^{1/p}.$$

Properties of these measures and the fuzzy integrals generated from them were studied. Keller and Yan [18, 38] have extensively studied the use of non-normalized possibility measures in fuzzy integral applications of multicriteria decision making and image segmentation.

We can also generalize the definition of the fuzzy integral [17, 24, 31, 33, 34] as a tool for information fusion. The generalization, originally proposed by Weber [33] and extensively studied in [24], involves replacing the minimum and maximum operators with more general aggregation operators, t-norms and t-conorms, resulting in increased flexibility. The original fuzzy integral can be interpreted as the "highest pessimistic" grade of agreement between the objective evidence and the subjective information embedded in the measure. By replacing the minimum in the definition of the fuzzy integral with an arbitrary t-norm, we can achieve an even more pessimistic evaluation, as might be desired when one wants to be very cautious. Let T be a t-norm, then this generalized fuzzy integral is given by

$$e_T = \bigvee_i \left[T(h(x_i), g(\{x_1, \ldots, x_i\}))\right].$$

Weber [33] has provided a counterexample that the generalization of the fuzzy integral when max is replaced by any t-conorm and min is replaced by any t-norm is not always well defined. In order to guarantee that the generalized integral is well defined, the operators must satisfy the distributive law [20]. In our work, we used the following t-norms, all of which are distributive with respect to maximum [24]:

Drastic product:

$$T_D(a,b) = \begin{cases} a, & b = 1, \\ b, & a = 1, \\ 0, & \text{else}. \end{cases}$$

Bounded product:

$$T_B(a, b) = \max(0, a + b - 1).$$

Algebraic product:

$$T_\Pi(a, b) = ab.$$

Logical product:

$$T_M(a, b) = \min(a, b).$$

It is easy to show that $T_D \leqslant T_B \leqslant T_\Pi \leqslant T_M$, and so, the resulting generalized fuzzy integrals maintain that same ordering [24].

In [26], Wierzchon introduced an "optimistic" version of the fuzzy integral by interchanging the order of the max and min operators in the original fuzzy integral equation. This also can be generalized by replacing the maximum operator in that formulation by a more optimistic t-conorm, S, to obtain:

$$E_S = \bigwedge_i \{S[h(x_i), g(\{x_1, \ldots, x_i\})]\}$$

for the finite case, where once again, the set X has been reordered so that $h(x_1) \geqslant h(x_2) \geqslant \cdots \geqslant h(x_n)$. Here, the generalized integral ranges from the "lowest optimistic" to "highest optimistic" combination of objective evidence and subjective information. This formulation was actually shown to be linearly related to the Weber generalized integral in [11], and hence, is not actually required to obtain the complete range of generalizations. However, it has an intuitive interpretation, and so, we keep its formulation and refer to both new integrals as generalized fuzzy integrals. As with the previous generalization, we used t-conorms for our applications which are distributive with respect to minimum. They included:

Drastic sum:

$$S_D(a, b) = \begin{cases} a, & b = 0 \\ b, & a = 0 \\ 1, & \text{else}. \end{cases}$$

Bounded sum:

$$S_B(a, b) = \min(1, a + b).$$

Algebraic sum:

$$S_\Pi(a, b) = a + b - ab.$$

Logical sum:

$$S_M(a, b) = \max(a, b).$$

It is easy to show that $S_M \leqslant S_\Pi \leqslant S_B \leqslant S_D$, and so, the resulting generalized fuzzy integrals maintain that same ordering.

As a comparison, we applied the generalized fuzzy integral pattern recognition algorithms to the ATR problem given above. The partial evaluation functions $h_i(x_j)$ for each class ω_i and each information source (feature type, here) x_j were computed as in the original problem. The densities were computed from the overlap of the feature histograms as in [16, 31] so that they were quite similar to those in Table 1. In fact, each density was computed as the fuzzy set theoretic complement of the ratio of the area of the intersection to the area of the union of the training data feature histograms for the two classes. To use an S-decomposable measure, the density set needed to be normalized. The original and the normalized densities are shown in Table 2.

We conducted several experiments with generalized fuzzy integrals with this data. In the results reported here, we used the Schweizer and Sklar based measure and computed various generalized fuzzy integral pattern recognition algorithms for this situation. We chose the Schweizer and Sklar measures since one of the goals in [31] was to determine the effect of the parameter in a family of t-conorms on the fuzzy integral results. The best overall results were obtained with e_D, i.e., the generalization obtained by replacing minimum with the drastic product. These results are displayed in Table 3. In this data set, many of the objects are easily distinguished using any approach. (This is usually the case in any pattern recognition problem.) The "difficult" objects in the test data here had the property that the true class was often best supported by the least important features (as determined by the training data). Hence, the most pessimistic generalized integral produced the optimal decision.

Table 2
Original and normalized densities for the generalized fuzzy integral

	g^1	g^2	g^3	g^4
Original densities				
Tank	0.11	0.26	0.19	0.24
APC	0.08	0.28	0.19	0.26
Normalized densities				
Tank	0.42	1.00	0.73	0.92
APC	0.29	1.00	0.68	0.93

Table 3
Results for the generalized fuzzy integral classifier with Min replaced by drastic product, i.e., e_D (*Confusion matrix*) (*Total correct* 93.4%)

	Tank	APC
Tank	176	0
APC	16	50

An interesting note is that since one can construct a separate integral for each class, it is possible to intentionally bias the results in favor of a particular class. This is similar to what is done by adding a loss function to the Bayes Decision Theory formulation. For example, Table 3 clearly demonstrates the importance of the class "Tank" in the recognition processes, that is, true-negative errors were minimized – no Tanks were misclassified as APCs. If the converse situation were crucial, i.e., classification of APCs was considered more important than that for Tank, the optimistic generalized integral formed by replacing maximum by the probabilistic sum S_Π and thus computing E_Π produces the results in Table 4. The point that we are

Table 4
Results for the optimistic generalized fuzzy integral classifier with Max replaced by probabilistic product, i.e., E_{Π} (Maximizes APC classification rate) (*Confusion matrix*) (*Total correct* 77.7%)

	Tank	APC
Tank	125	51
APC	3	63

making here is that the fuzzy integral methodology is capable of supplying extremely flexible operators for the fusion of information in pattern recognition. Of course, some form of supervision is needed to determine the appropriate choices for a particular problem domain. If the testing data is always similar to the training data, then automated optimization techniques could be employed to choose the correct parameters for the final algorithm. However, if the test set shows significant differences from the training (as was the case in the ATR problem), then some form of human intervention is necessary.

In the methods discussed above, one needs to compute membership values in different classes from observed feature data. Several methods can be used for this purpose. One approach is to run the FCM algorithm on the training data to estimate the prototypes which can then be used to compute membership values, i.e., the objective evidence $h_i(x_j)$. Recently, we have used normalized histograms of the feature values generated from training data to estimate the particular membership functions [30, 31]. This has the advantages that it does not force any particular shape to the resultant distributions, can be extended to deal with multiple features instead of gray level alone, and can easily accommodate the addition of new classes. Krishnapuram and Keller [22, 23] have introduced a new class of unconstrained fuzzy clustering algorithms, called possibilistic clustering, which is robust in the presence of noise. In this context, they have the advantage that the final memberships for each cluster are decoupled from the rest, providing a good membership function for the above recognition techniques.

4. Fusion of classifiers with the fuzzy integral

In this section, we describe experiments in which the fuzzy integral was used to fuse the results of separate classifiers. Two applications are considered. The first example involves an extension of the work in automatic target recognition discussed above to a multisensor recognition problem. The second example involves a new approach to handwritten character classification by fusing classification information from two independently trained classifiers.

4.1. Multisensor fusion with the generalized fuzzy integral

As mentioned earlier, the information sources for fuzzy integration can, in fact, be classifiers themselves. One example of such fusion was reported in [30]. Here, we consider the case of using generalized fuzzy integrals as partial classifiers (on different feature sets) and then fusing that output with a second integration. Similar to the data described in Section 3, independent FLIR and TV sequences of images were used to build a training set of objects, and the algorithms were tested on registered FLIR and TV sequences. In this experiment, four statistical features and seven texture features were extracted from each image window under investigation for each sensor. There were 200 Tanks and 100 APC's in both the training and testing data sets. At the first level of integration, we combined the feature evidence of each type for each sensor, resulting in four classifiers: TV-Stat, TV-Texture, FLIR-Stat, and FLIR-Texture. The outputs of these classifiers were then combined at the top level to produce a final confidence value for each class. The fuzzy densities at the first level (features) were generated as in the earlier experiments from overlap ratios of training data histograms, whereas the densities for the final integration were determined by measuring how well each ensemble of features performed in a 10% Jackknife experiment using a multivariate Bayes classifier, i.e., the densities were the normalized percentage of correct classification averaged over the 10 trials.

The generalized fuzzy integrals defined in Section 3 were all used at both stages of the classification

Table 5
Confusion matrix for the multisensor ATR problem using E_Π at the feature level and e_M at the classifier fusion level (*Confusion matrix*) (*Total correct* 97.0%)

	Tank	APC
Tank	195	5
APC	4	96

process. In this experiment, the test data was considerably different in feature space from the training (it was sleeting when the test data was collected). As a result, values of the evidence functions $h_i(x_j)$ were quite low (essentially 0) for many of the image windows for several features, x_j. What we discovered is that for a situation like this, the optimistic version of the generalized fuzzy integral produces superior results. Hence, the best results were obtained using E_Π at the feature classification level and then follow that with a more restrictive integral, e_M at the classifier fusion level. Table 5 shows the confusion matrix for this choice. As can be seen, even though the test data differed significantly from the training data, the multistage generalized fuzzy integral provided excellent classification results.

4.2. Fusion of handwritten character classifiers

Classifiers trained to recognize handwritten characters can also be combined using the fuzzy integral. In this example, two high-performance neural network based classifiers are combined in this way. Classification is performed on upper case characters and lower case characters separately. There are two classifiers for each case and each classifier has 26 output units, one for each class. The two classifiers are multi-layer feedforward neural networks trained using standard backpropagation. The target outputs are set using the fuzzy k-nearest neighbor algorithm [15] in order to more accurately represent the inherent ambiguity in character classification [5–9]. The two classifiers use different features as inputs, the bar-features and transition features [8, 10]. The classification rates

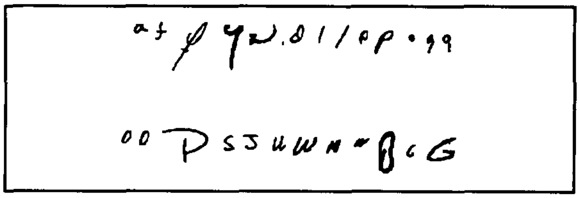

Fig. 1. Samples of characters from handwritten character data.

obtained using the fuzzy integral are significantly higher than those obtained using the individual classifiers.

The data sets used for the experiment consist of isolated handwritten characters extracted from images of addresses from United States Postal Service (USPS) mail. Due to the distribution of characters in the USPS mailstream, a variable number of characters were available from each class. Balanced training and testing sets were constructed consisting of 250 characters per class for each of the training and testing sets. Most classes had sufficient characters available to construct the training and testing sets. For those classes with fewer than 500 samples, images of available characters were resampled to randomly chosen sizes and used as different samples. A sample of some of the characters is shown in Fig. 1. As can be seen, these date sets are very challenging.

Since there are 26 output units for each neural network classifier, there are 26 fuzzy measures to estimate for each case. The densities for each class are estimated using the normalized confusion matrix. The normalized confusion matrix for a classifier is a matrix $\boldsymbol{P} = (p_{ij})$ where p_{ij} is the ratio of the number of samples in the training set from class i that are assigned to class j to the total number of samples in the training set from class i. The assignment rule is to assign a sample to the class with the highest output activation value. The densities are estimated using the formula:

$$g^i = \left(\frac{1}{n-1} \sum_{k \neq i} (1 - p_{ki}) \right) p_{ii}$$

where n is the number of classes. The density from class i is the product of the percentage of the samples from class i that are assigned to class i and a factor which decreases as the number of samples

Table 6
Density values for character recognition fuzzy integrals

Classes	bf-lc	bf-uc	tf-lc	tf-uc
a, A	0.742	0.890	0.671	0.913
b, B	0.833	0.878	0.715	0.854
c, C	0.842	0.886	0.729	0.862
d, D	0.865	0.743	0.779	0.704
e, E	0.678	0.876	0.534	0.832
f, F	0.862	0.812	0.657	0.802
g, G	0.703	0.891	0.801	0.820
h, H	0.654	0.864	0.656	0.835
i, I	0.727	0.778	0.692	0.735
j, J	0.959	0.855	0.945	0.840
k, K	0.675	0.857	0.672	0.763
l, L	0.775	0.748	0.600	0.816
m, M	0.869	0.875	0.801	0.865
n, N	0.633	0.757	0.698	0.755
o, O	0.837	0.857	0.758	0.832
p, P	0.926	0.869	0.853	0.853
q, Q	0.951	0.958	0.919	0.959
r, R	0.751	0.831	0.718	0.804
s, S	0.763	0.808	0.512	0.812
t, T	0.748	0.810	0.734	0.759
u, U	0.776	0.901	0.686	0.805
v, V	0.817	0.728	0.698	0.719
w, W	0.853	0.762	0.884	0.825
x, X	0.797	0.928	0.719	0.936
y, Y	0.858	0.868	0.779	0.819
z, Z	0.899	0.960	0.843	0.954

from other classes that are classified as class i increases. Thus, the densities measure the reliability of each information source. The densities are shown in Table 6. The headings bf-lc, bf-uc, tf-lc, and tf-uc indicate the feature type (bar or transition feature) followed by the case (upper or lower case).

In the special case of two inputs we have that the fuzzy integral output confidence for class i is given by the following rule: Let $b(i)$ and $t(i)$ denote the output activation for class i for the bar and transition feature based neural networks respectively and let

$$h_{\max} = \max(b(i), t(i)),$$

$$h_{\min} = \min(b(i), t(i)),$$

$$g_{\max} = g_b^i \text{ if } h_{\max} = b(i),$$
$$= g_t^i \text{ else.}$$

Table 7
Comparison of character recognition rates using fuzzy integral to combine classifiers

Network	UC(Train)	UC (Test)	lc (Train)	lc (Test)
Bar-Feature	88.62%	85.71%	83.88%	79.92%
Transition-Feature	87.14%	82.91%	84.15%	80.15%
Fuzzy integral	90.58%	89.38%	86.60%	84.08%

Then

$$\text{Fuzzy\_integral\_conf}(i) = \vee\ [h_{\max} \wedge g_{\max}, h_{\min}].$$

The classification results are given in Table 7. The fuzzy integral combination scheme achieves a significantly higher scores than each individual classifier. A difference of 4% in classification rates scores on isolated characters can result in significant increase in classification rates for strings of characters. For example, assuming independence, the classification rate for strings of length n given an isolated character rate of r is r^n.

5. Conclusion

In this paper, we examined the use of the fuzzy integral and its generalizations as tools for pattern recognition. We demonstrated that the fuzzy integral can be used to achieve significantly higher classification rates on nontrivial real world applications involving automatic target recognition and handwritten character recognition. These demonstrations involve using the fuzzy integral at several levels: combination of feature information, combination of sensor-based information, and fusion of classifier information. The core problem in using these tools effectively is estimating the densities. We have discussed several methods for density estimation in the context of the above applications. Density or fuzzy measure estimation remains an active area of research. Our work and that of others shows that the fuzzy integral and its generalizations can achieve increased classification rates in pattern recognition problems. The full potential of the technique is as yet undetermined.

References

[1] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms* (Plenum Press, New York, 1981).

[2] J. Bezdek, Ed., *Proceedings of the First IEEE Conference on Fuzzy Systems*, San Diego, CA (1992).

[3] J.C. Bezdek and S.K. Pal, Eds., *Fuzzy Models for Pattern Recognition* (IEEE Press, New York, 1992).

[4] D. Dubois and H. Prade, A class of fuzzy measures based on triangular norms, *Internat. J. General Systems* **8** (1982) 43–61.

[5] P.D. Gader, M. Mohamed and J. Chiang, Comparison of crisp and fuzzy character networks in handwritten word recognition, *Proc. North American Fuzzy Inform. Process. Soc.*, Puerto Vallarta, Mexico (1992) 257–266.

[6] P.D. Gader, M. Mohamed and J. Chiang, Segmentation-based handprinted word recognition, *Proc. 5th United States Postal Service Advanced Technology Conference*, Washington, D.C. (1992) 215–225.

[7] P.D. Gader, M. Mohamed and J. Chiang, Fuzzy and crisp handwritten alphabetic character recognition using neural networks, *Proc. of Artificial Neural Networks in Engineering*, St. Louis, MO (1992) 421–426.

[8] P.D. Gader, M. Mohamed and J. Chiang, Comparison of crisp and fuzzy character neural networks in handwritten word recognition, submitted to *IEEE Trans. on Fuzzy Systems* (1993).

[9] P.D. Gader, A.M. Gillies and D. Hepp, Handwritten character recognition, in: E. Dougherty, Ed., *Digital Image Processing: Fundamentals and Applications* (Marcel Dekker, New York, 1994).

[10] P.D. Gader, J.M. Keller and J. Cai, A fuzzy logic system for the detection and recognition of street number fields on handwritten postal addresses, accepted for publication in *IEEE Trans. on Fuzzy Systems* (1993).

[11] M. Grabisch and M. Sugeno, Fuzzy integral and dual measures: applications to pattern recognition, *Sino-Japan Symposium on Fuzzy Sets and Systems*, Beijing, China (1990).

[12] M. Grabisch, T. Murofushi and M. Sugeno, Fuzzy measure of fuzzy events defined by fuzzy integrals, *Fuzzy Sets and Systems* **50** (1992) 293–313.

[13] M. Grabisch and M. Sugeno, Multi-attribute classification using fuzzy integral, *Proc. First IEEE Conference on Fuzzy Systems*, San Diego, CA (1992) 47–54.

[14] M. Grabisch and F. Dispot, A comparison of some methods of fuzzy classification on real data, *Proc. 2nd International Conference on Fuzzy Logic and Neural Networks*, Iizuka, Japan (1992) 659–662.

[15] J. Keller, M. Gray and J. Givens, A fuzzy k-nearest neighbor algorithm, *IEEE Trans. on System, Man, and Cybernetics* **15** (1985) 580–585.

[16] J. Keller, H. Qiu and H. Tahani, The fuzzy integral in image segmentation, *Proc. NAFIPS-86*, New Orleans (1986) 324–338.

[17] J. Keller and H. Tahani, The fusion of information via fuzzy integration, *Proc. NAFIPS '92*, Puerto Vallarta, Mexico, December (1992) 468–477.

[18] J. Keller and B. Yan, Possibility expectation and its decision making algorithm, *Proc. First IEEE Conference on Fuzzy Systems*, San Diego, CA (1992) 661–668.

[19] J. Keller and R. Krishnapuram, Fuzzy decision models in computer vision, in: R. Yager and L. Zadeh, Eds., *Fuzzy Sets, Neural Networks, and Soft Computing*, (Van Nostrand, New York, 1994) pp. 213–232.

[20] G. Klir and T. Folger, *Fuzzy Sets, Uncertainty, and Information* (Prentice Hall, Englewood Cliffs, NJ, 1988).

[21] B. Kosko, *Neural Networks and Fuzzy Systems* (Prentice Hall, Englewood Cliffs, NJ, 1992).

[22] R. Krishnapuram and J. Keller, A possibilistic approach to clustering, *IEEE Trans. on Fuzzy Systems*, **1**(2) (1993) 98–110.

[23] R. Krishnapuram and J. Keller, Fuzzy and possibilistic clustering methods for computer vision, in: S. Mitra, M. Gupta and W. Kraske, Eds., *Neural and Fuzzy Systems* (SPIE Press, 1994) pp. 133–159.

[24] T. Murofushi and M. Sugeno, Fuzzy t-conorm integral with respect to fuzzy measures: generalization of Sugeno integral and Choquet integral, *Fuzzy Sets and Systems* (1991) 57–71.

[25] J. Osborn, Training and fuzzy integral, MS Thesis, University of Missouri-Columbia (1992).

[26] H. Qiu and J. Keller, Multispectral segmentation using fuzzy techniques, *Proc. NAFIPS-87*, Purdue University (1987) 374–387.

[27] G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, 1976).

[28] M. Sugeno, Fuzzy measures and fuzzy integrals: A survey, in *Fuzzy Automata and Decision Processes* (North Holland, Amsterdam, 1977) 89–102.

[29] M. Sugeno and T. Murofushi, Choquet integral as an integral form for a general class of fuzzy measures, *Proc. 2nd IFSA Congress*, Tokyo, Japan (1987) 408–411.

[30] H. Tahani and J. Keller, Information fusion in computer vision using the fuzzy integral, *IEEE Trans. on System, Man and Cybernetics* **20**(3) (1990) 733–741.

[31] H. Tahani, The generalized fuzzy integral in computer vision, Ph.D. Dissertation, University of Missouri-Columbia (1990).

[32] Z. Wang and G. Klir, *Fuzzy Measure Theory* (Plenum Press, New York, 1992).

[33] S. Weber, S decomposable measures and integrals for archimedean t-conorms S, *J. Math Analysis and Applications* **101** (1984) 114–138.

[34] S. Wierzchon, On fuzzy measure and fuzzy integral, in: M. Gupta and E. Sanches, Eds., *Fuzzy Information and Decision Process* (North Holland, Amsterdam, 1982) 79–86.

[35] J. Wootton, J. Keller, C. Carpenter and G. Hobson, A multiple hypothesis rule-based automatic target recognizer, in: J. Kittler, Ed., *Pattern Recognition*, Lecture Notes in Computer Science, Vol. 301 (Springer, Berlin, 1988) 315–324.

[36] R.R. Yager and L.A. Zadeh, Eds., *An Introduction to Fuzzy Logic Applications in Intelligent Systems* (Kluwer Academic, Norwell, MA, 1992).
[37] B. Yan and J. Keller, Conditional fuzzy measures and image segmentation, *Proc. NAFIPS-91*. Columbia, MO (1991) 32–36.
[38] B. Yan and J. Keller, A possibility integral-based image segmentation approach, *IEEE Trans. on Fuzzy Systems*, submitted (1993).
[39] H.-J. Zimmermann, *Fuzzy Set Theory and Applications* (Kluwer-Nijhoff, Boston, 1991).