Improving Fuzzy Rule Classifier by Extracting Suitable Features From Capacities With Respect to the Choquet Integral

Emmanuel Schmitt, Vincent Bombardier, and Laurent Wendling

Abstract—In this paper, an iterative method to select suitable features in an industrial pattern recognition context is proposed. It combines a global method of feature selection and a fuzzy linguistic rule classifier. It is applied to an industrial fabric textile context. The aim of the global vision system is to identify textile fabric defects. From the related industrial process, the training data sets are small, and some are incomplete. Moreover, the recognition step must be compatible with the time constant of the system, which generally imposes low complexity for the system. The choice of the most relevant features and the reduction of their number are important to respect these constraints. The feature selection method is based on the analysis of indexes extracted on the lattice defined from training in relation with the Choquet integral. This selection step is embedded in an iterative algorithm to discard weaker features in order to decrease the number of rules while keeping good recognition rates. The recognition step is done with a fuzzy reasoning classifier that is well adapted for this application case. The proposed method is quite efficient with small learning data sets because of the generalization capacity of both the feature selection and recognition steps. The experimental study shows the wanted behavior of this approach: the feature number decreases, whereas the recognition rate increases. Thus, the total number of generated fuzzy rules is reduced.

Index Terms—Choquet integral, feature selection, fuzzy logic, image processing, pattern recognition.

I. INTRODUCTION

THE FEATURE selection method proposed in this paper takes place in a problematic of complexity reduction. The application domain relates to quality control in a technical textile industry (carbon fibers, Kevlar, etc.). The aim of the global vision system is to identify fabric textile defects in a continuous mode during production.

E. Schmitt is with the Centre de Recherche en Automatique de Nancy, CNRS UMR 7039, Université Henri Poincaré—Campus Scientifique, 54506 Vandoeuvre-lès-Nancy Cedex, France (e-mail: emmanuel.schmitt@cran.uhp-nancy.fr).

V. Bombardier is with École Supérieure d'Informatique et Applications de Lorraine and Centre de Recherche en Automatique de Nancy, CNRS UMR 7039, Université Henri Poincaré—Campus Scientifique, 54506 Vandoeuvre-lès-Nancy Cedex, France (e-mail: vincent.bombardier@cran.uhp-nancy.fr).

L. Wendling is with the École Supérieure d'Informatique et Applications de Lorraine and Laboratoire Lorrain de Recherche en Informatique et ses Applications, UMR 7053, Université Henri Poincaré—Campus Scientifique, 54506 Vandoeuvre-lès-Nancy Cedex, France (e-mail: laurent.wendling@loria.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCB.2008.925750

Because of this specific industrial context, there are many constraints. One constraint is the necessity of working with very small training data sets (sometimes there are only one or two samples for a class defect because of its rareness). Another difficulty is to respect the real-time constraints in the industrial production system. Therefore, low complexity must be kept for the recognition model. Thus, this paper takes place on a "small scale" domain according to [1] and [2] definitions because of the weak number of used features.

Moreover, defects are intrinsically fuzzy. For example, there is not always a strict boundary between a "sound" fabric and a "defective" region; this transition is more or less gradual. The segmentation step provides an accurate "defective" region, i.e., the achieved features are accurate but uncertain (the same defect could be processed twice without providing the same characteristic vector).

Thus, the recognition method must take these specificities into account. Using fuzzy logic minimizes this effect to obtain a measure that is less sensible to these uncertainties. Therefore, a classifier based on fuzzy linguistic rules has been chosen.

The second part of this paper focuses on the selection of suitable parameters in order to decrease the number of rules. Handling with several classifiers allows integration of their discriminatory aspects to improve the recognition step [3]. Despite pattern recognition, methods are generally independently built. Their combination may lead to positive correlations, because both aim at achieving the same goal, and both are based on the same learning data [4], [5]. Nonetheless, even if approaches like Adaboost, arcing [6], and boosting [7] try to limit this dependence by reinforcing diversity, it is difficult to measure it in order to efficiently incorporate it into the classification process [3], [8].

Furthermore, such methods often require a consistent amount of learning data to be efficient [9]. Many classifier combination systems have been proposed and compared in the literature [10]–[13], and a full presentation of most of these could be found in a reference book by Duda *et al.* [9]. In the application-studied context, only two sets of learning data and simple parameters are processed while considering processing time constraint is handled. Furthermore, data may be inconsistent due to the fast acquisition step.

The proposed method consists in analyzing the learning database used to generate the rule-based system. Over the years, many aggregation operators were introduced. For most of these operators, the relative importance of a decision criterion (DC) in the final decision is represented by a weight assigned to it.
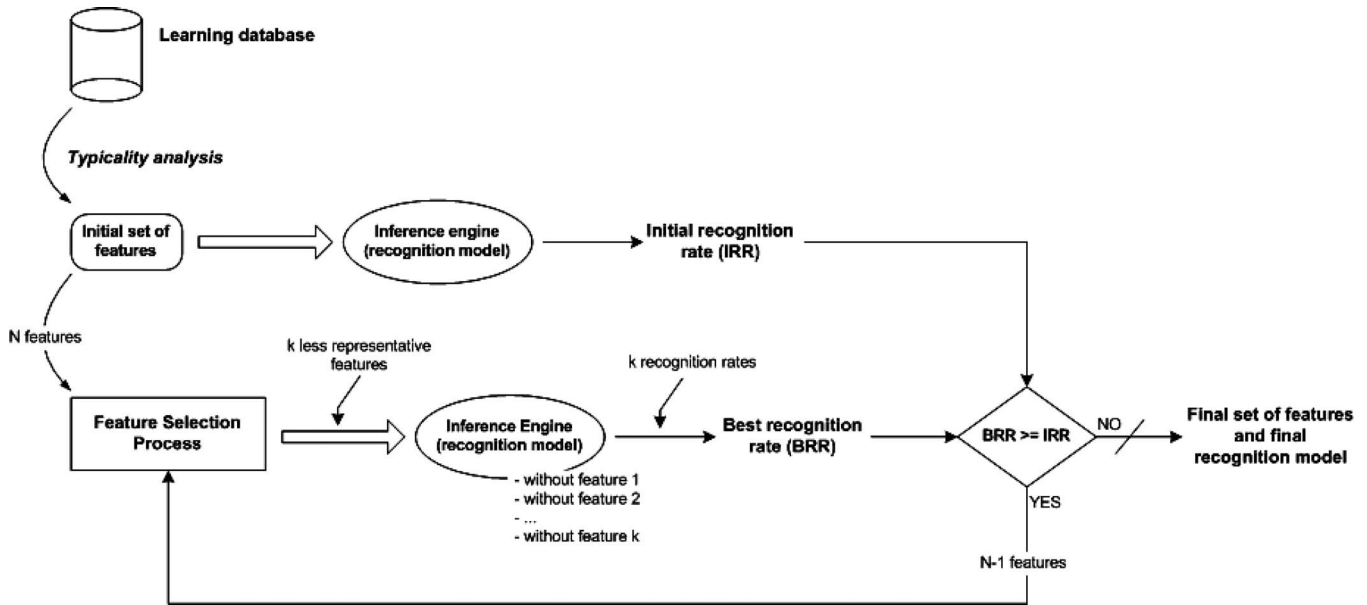
Fig. 1.   Overall description of the recognition system.

Well-known operators applying this scheme include the quasi-arithmetic means, the weighted minimum and maximum, and the ordered weighted averaging [14]. None of these operator families, however, take the possible interactions between the constituents of the aggregation into account. The Choquet integral is then considered because it allows taking into account of such interactions while generalizing many aggregation operators by choosing specific fuzzy measure such as weighted arithmetic means, ordered weighted averages, order statistics, and median [15], [16].

Fuzzy integrals, and the Choquet integral in particular, have been successfully used as fusion operators in various applications [17], [18] including content-based image retrieval [19] and speech recognition [20].

A fuzzy measure learning scheme is used with respect to the Choquet integral as it allows consistent learning even if only a few samples per class are processed. Such an algorithm is suitable for this specific application where few learning data are provided by the industry and also to handle ambiguous features. As numerous rules are provided by the system, it is neither easy to determine which features are important nor easy to determine which features are redundant.

Two importance and interaction indexes can be calculated on fuzzy measures. These indexes are used here to define a set of weaker features to be removed from the initial set. An iterative algorithm combining a fuzzy rule classifier (FRC) and suitable feature extraction method is proposed in order to decrease the number of rules while keeping good recognition rates. Both parts of this algorithm have been checked separately in other contexts (respectively [21] and [22]). They have proven their efficiency in "small scale" contexts with small learning data sets. The two approaches are merged in order to enhance the FRC by selecting suitable features.

This method has been applied on five academic University of California—Irvine (UCI) [23] and three industrial ones. They are used to enhance the proposed method by showing some

interesting results such as a large decrease in number and an improvement in recognition rates.

A graphical description of the main components of the system is shown Fig. 1.

The outline of this paper is as follows. In Section II, some background on the fuzzy linguistic rule classifier and its use in a fuzzy recognition system are introduced. In Section III, a scheme to discard weaker parameters using the Choquet integral is given. Section IV presents the global method and its application in an industrial context. A study of the behavior of the proposed model using well-known databases is also provided to show the portability of the system.

II. FUZZY LINGUISTIC RULE CLASSIFIER

The fuzzy reasoning classifier [24] is based on a fuzzy linguistic rule mechanism. It is well adapted to the presented industrial application. Indeed, it presents a very good and efficient generalization from a few sample sets and is able to provide gradual membership for output classes [21]. Its satisfactory behavior has been shown in [21] by several comparisons with other classifiers such as k-NN, neural networks, or support vector machine (SVM).

This implemented algorithm, for the fuzzy recognition method, is a supervised learning mechanism divided into three parts:

1) input fuzzification (features of the characteristic vector);
2) fuzzy rule generation;
3) rule adjustment.

Then, the obtained rule set is used by the generalization step, and the output class is determined by the rule of maximal answer. Fig. 2 shows the different steps of the fuzzy recognition method.

It must be noted that there is no defuzzification step in the proposed method, because the exposed problem is quite
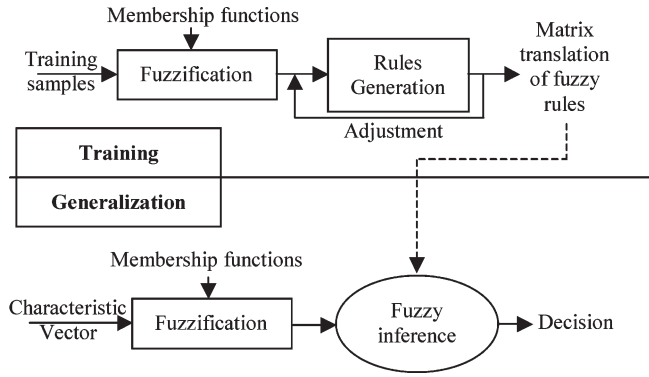
Fig. 2.   Overall description of the fuzzy recognition system.

Fig. 3.   Example of fuzzification of the parameter "Size" into three subsets.

different from the fuzzy control application. Symbolical outputs are needed, not numerical ones. This output formalism contributes to reducing the difference existing between both industrial textile and image processing vocabularies. This problem is often referred to as the "semantic gap," defined as "the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation [25]."

The chosen formalism also improves the interpretability of the system.

A. Input Fuzzification Step

The fuzzification step aims to translate numerical variables into linguistic variables. A linguistic variable [26], [27] is defined by a triple value (V, X, Tv), where V, X, and Tv are defined as follows.

1) V is a variable (area, size, etc.) defined on a set of reference X.
2) X is the universe of discourse (field of variation of V).
3) Tv is the chosen vocabulary to describe in a symbolic way the values of V (small, big, dark, light, etc.).

The set $\text{Tv} = \{A_1, A_2, \dots, \}$, finite or infinite, contains normalized fuzzy subsets of X which are usable to characterize V. Each fuzzy subset Ai is defined by the membership degree $\mu_{\text{Ai}}(x)$.

This fuzzification step defines the decomposition number of the considered variable to provide the fuzzy rule premises.

For example, the membership function for variable V, called "Size," is initialized with respect to the data analysis of the training sample set. The symbolic vocabulary then associated with this variable V is $\text{Tv} = \{\text{Small}, \text{Medium}, \text{Big}\}$. Therefore, the linguistic variable "Size" is split into three terms, and this variable is characterized by a vector composed of three membership degrees: $[\mu_{\text{Small}}(x), \mu_{\text{Medium}}(x), \mu_{\text{Big}}(x)]^{\text{T}}$ as shown in Fig. 3.

To summarize this step, one characteristic is represented by distributed terms in its definition field, called the universe of discourse, according to its useful and variable parts.

The different terms are chosen in relation to the expert vocabulary. The number of terms used to qualify a linguistic variable
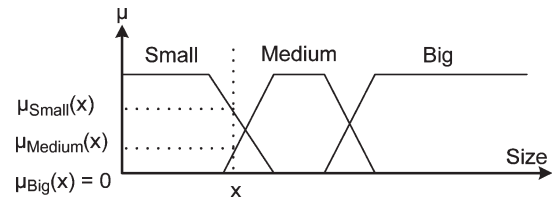
is, generally, empirically defined. However, the industrial user, who is not an expert in pattern recognition, often chooses a regular distribution of terms, generally having more terms than are needed. Whenever the number of terms increases, so does the number of rules and, thus, the overall complexity of the entire system.

An automatic fuzzification method can also be used. Classical automatic methods are based on a genetic algorithm [28] or clustering [29], [30]. However, these kinds of methods need large number of training samples to succeed. Moreover, if the partition of the variable input space is not fit with real data, the terms will be inappropriate.

The chosen fuzzification method is based on the study of the output class typicality. For example, the typicality measure $T(V)$ is computed with (1) from extern dissimilarity and intern likeness according to the output classes [31]

$$T(x_a^u) = \frac{R \cdot D}{R \cdot D + (1 - R) \cdot (1 - D)} \tag{1}$$

$$R(x_a^u) = \frac{\sum\limits_{i=1}^{n} \frac{1}{d\left(x_a^u, x_a^{f_i}\right)}}{n} \tag{2}$$

$$D(x_a^u) = \frac{\sum\limits_{i=1}^{m} 1 - d\left(x_a^u, x_a^{e_i}\right)}{m} \tag{3}$$

where

x_a^u value of parameter a for sample x;

$x_a^{f_i}$ value of parameter a for sample f belonging to the same class as x;

$x_a^{e_i}$ value of feature a for sample e not belonging to the same class as x;

$d(x, y)$ Euclidian distance;

n number of samples which belong to the same class as sample x;

m number of samples which do not belong to the same class as sample x.

From the typicality measure $T(V)$, the correlation (Corr), and the cross correlation (Xcorr), coefficients are computed for each output class. Then, from the ratio Corr/Xcorr, which characterizes interclass similarities, the number of terms is determined. Their positions are obtained by calculating the mean value of the samples belonging to the considered output classes [21].

The main interest takes place in the automatic adaptation of the fuzzification step which makes the tuning of the system easier.

B. Fuzzy Rule Generation

This second step allows definition of "If ... Then ..." fuzzy rules, e.g.,

≪ IF the Height is (High) AND the Length is (Small)

THEN the defect IS (Inclusion) ≫

Each rule describes the perceived defect related to the system. Such rules can be classified into two categories: conjunctive rules and implicative rules. These two categories are regrouped, respectively. On the one hand, there are the possibility rules and the antigradual rules and, on the other hand, the certitude rules and the gradual rules [32]. The conjunctive rules are derived from the data analysis field where reasoning mechanisms are led by the data, whereas implicative rules are most utilized in the cognitive sciences field where reasoning is led by knowledge [33].

For this application, conjunctive reasoning mechanisms have logically been selected. Each rule is activated in parallel, and a disjunction operator combines the intermediate results. This inference mechanism gives interpretation and semantics, which differ from mechanisms using implications [32]. In particular, it assures the consistency of the rule base [34]. If no information is processed, i.e., the input space is not covered by the rule set, the output gives an "unknown defect" class. The two main models using these rules are the Larsen's model and the Mamdani's model [35]. The Sugeno's model [36] is not suitable in this case because the aim is not to achieve numerical output values.

The chosen classifier is based on Ishibuchi's algorithm which provides an automatic rule generation step [37]. There are many methods which automatically obtain fuzzy rules according to data sets such as a genetic algorithm [38]–[40] or the decision tree method (DTM) [41], [42], but the Ishibuchi's algorithm is quite simple and gives better results [21]. Moreover, its inference mechanism follows the Larsen's model, which is better than the Mamdani's model, because the product is more adapted than the minimum for the manipulation of several premises [40]. In fact, it allows nonlinear splitting of variable input space. The iterative version of the Ishibuchi's algorithm [43] is used here. It allows one to adjust the input space splitting by supporting the rule of having the maximum response.

The expert must prepare defective sample sets to generate fuzzy rules via an automatic rule generation algorithm [37], [43].

If two linguistic variables are considered for input (V_1, V_2) and one for output (Z_3), the general form of the associated fuzzy rule is [44]

$$\text{if } V_1 \text{ is } A_i \text{ and } V_2 \text{ is } A_j, \text{ then } Z \text{ is } C_k \tag{4}$$

V_1, V_2 input linguistic variables defined on X_1 and X_2 (intensity and surface, for example);

Z_3 output linguistic variable defined on Y (defect name);

A_i and $A_j \in$ Tv   (chosen vocabulary "light" and "dark," for example);

C_k class of kth defect.

This rule defines an implication, where the antecedent (premiss) is

$$V_1 \text{ is } A_i \text{ and } V_2 \text{ is } A_j \tag{5}$$

and the consequent is

$$Z \text{ is } C_k. \tag{6}$$

In (5), the "and" operator corresponds to the Cartesian product between V_1 and V_2 linguistic variables [35]. This operation is done with a T-Norm. A product is used here

$$T(x_1, x_2) = \mu_A(x_1) * \mu_B(x_2). \tag{7}$$

Then, the implication I is done through the Generalized Modus Ponens mechanism based on the use of maximum/product composition law. This inference follows Larsen's model [35], which uses a pseudoimplication operator represented by the product

$$I(V1, V2, Z) = T(T(x1, x2), y) \tag{8}$$

$$\beta_{Ck} = [\mu_A(x_1) * \mu_B(x_2)] * \mu_Z(y). \tag{9}$$

Finally, each rule gives a partial conclusion. β is aggregated to the others according to a fuzzy operator of disjunction. The disjunction operator is represented by the maximum operator according to Zadeh's case

$$\beta_{CX} = \max\{\beta_{C1}, \beta_{C2}, \ldots, \beta_{CM}\} \tag{10}$$

where β_{CX} corresponds to the maximum membership degree given by the rule defined on $A_i \times A_j$.

Then, the CF confident coefficient is calculated from the preceding truth degree

$$\text{CF}_{ij} = (\beta_{CX} - \beta) / \sum_{T=1}^{M} \beta_{CT} \tag{11}$$

where

$$\beta = \sum_{\substack{T=1 \\ CT \neq CX}}^{M} \beta_{CT} / (M-1)$$

and

$$\beta_{CT} = \sum_{x \in CT} \mu_i(x_1) \times \mu_j(x_2).$$

C. Rule Adjustment

The adjustment represents the iterative part of the algorithm. The following mechanism allows for the adjustment of the decomposition of representative space according to the achieved results [32].

1) From the training patterns, the algorithm generates the first model.
2) If the classification rate is below a ε threshold, defined by the user, the iterative part is performed to adjust this rate.

In fact, fuzzy rules are generated again by both injecting the training patterns and considering the new response of each rule and adjusting the $\mathrm{CF}_{ij}^{I,J}$ confident coefficient with the following equations.

When x is properly classified by the R_{ij} rule, the adjustment of the CF confident coefficient is done by

$$\mathrm{CF}_{ij} = \mathrm{CF}_{ij} + \eta_1(1 - \mathrm{CF}_{ij}). \tag{12}$$

On the opposite side, when x is poorly classified by the R_{ij} rule, the adjustment of the CF confident coefficient is done by

$$\mathrm{CF}_{ij} = \mathrm{CF}_{ij} - (\eta_2 \times \mathrm{CF}_{ij}). \tag{13}$$

Tuning values have been chosen empirically in order to obtain the best recognition rates

$$\eta_1 = 0.4, \quad \eta_2 = 0.004, \quad \text{iteration number} = 500, \quad \varepsilon = 92\%.$$

The algorithm proposes an additional refining step. This step allows one to improve the membership degree of the maximum membership class by modifying the slope of its membership function. This way is not studied here due to the graduality of the answers that need to be kept because this vagueness improves the generalization capability of the classifier.

III. EXTRACTING SUITABLE FEATURES USING THE CHOQUET INTEGRAL

A. Choquet Integral

The Choquet integral was first introduced in capacity theory. Only useful definitions to explain the related work are presented here. In particular, only discrete spaces are considered. A thorough description of the Choquet integral is given in [15], [45], and [46].

Let m classes, $C_1, \ldots, C_m$, and n DC, $X = \{D_1, \ldots, D_n\}$ be considered. By DC, a feature description is considered, and an associated similarity ratio (min over max) is produced to ensure that any DC are in the same range, here [0, 1], and are able to be combined. Let x_0 be a pattern. For each DC, the aim is to calculate the confidence degree in the statement "According to D_j, x_0 belongs to the class C." Let P be the power of X, a capacity or fuzzy measure μ, defined on X; μ is a set function

$$\mu : P(X) \to [0, 1] \tag{14}$$

verifying the following axioms.
1) $\mu(\varnothing) = 0, \mu(X) = 1$.
2) $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$.

Fuzzy measures generalize additive measures by avoiding the additivity axiom. In this application context of the DC fusion, $\mu(A)$ represents the weight of importance or the degree of trust in the decision provided by the subset A of DC. The next step in building a final decision is to combine the Choquet integral with the partial confidence degree, according to each DC, into a global confidence degree.

Let μ be a fuzzy measure on X. The discrete Choquet integral of $\varphi = [\varphi 1, \ldots, \varphi n]^t$ with respect to μ, noted $C\mu(x)$,

is defined by

$$C_\mu(\varphi) = \sum_{j=1,n} \varphi(j) \left[\mu\left(A_{(j)}\right) - \mu\left(A_{(j+1)}\right) \right] \tag{15}$$

where $\varphi(1) \leq \ldots \leq \varphi(n)$. Also, $A(j) = \{(j), \ldots, (n)\}$ represents the $[j \ldots n]$ associated criteria in increasing order and $A(n+1) = \varnothing$.

B. Learning Data

The purpose of this step is to determine the more suitable learning data, taking into account the existing confusion between DC. A training pattern yields m training samples $\Phi_1, \ldots, \Phi_m$, with $\Phi_i = (\varphi_{i1}, \ldots, \varphi_{im})$ where φ_{ij} represents the confidence in the fact that the sample belongs to class i, according to DC j. For each of these samples, a target value must be assigned. For techniques using a different fuzzy measure per class, the optimal target value that minimizes the quadratic error is known [47]. A classical way to learn a single fuzzy measure, with respect to the Choquet integral, is to consider $C\mu(\Phi_i) = 1$, if the sample belongs to class i, and is otherwise zero. The larger the error between the real values of the Choquet integral is, the more the fuzzy measure coefficient will be modified. The Choquet integral, being an averaging operator output, values of zero or one will never be reached in real data. Hence, each learning sample will move the fuzzy measure away from the weighted arithmetic mean whether it is already correctly recognized or not. The proposed idea is to link the target value assigned to a sample with an estimation of the confusion between classes. That is an assessment of the degree for each pair of classes (C_i, C_j) that an element of C_i is recognized as belonging to C_j.

The confusion between classes is estimated by first building the confusion matrix for each DC. Then, an average confusion matrix is built by averaging theses matrices. Following the global confusion among classes, a decreasing function is defined to take it into account. The more important the confusion is, the closer to zero the value becomes. Thus, the target value for a sample, which is associated with the class having the least confusion, is the outcome of the Choquet integral. With such a target value, these samples leave the fuzzy measure unchanged when processed by the learning algorithm. On the opposite side, the target for the sample which is associated with the class having the most confusion is set to zero. This implies the biggest modification possible on the fuzzy measure.

C. Learning Step

The calculation of the Choquet integral requires the definition of the fuzzy measure, i.e., the assessment of any set of $P(X)$, which by definition is $\mu(\varnothing) = 0$, $\mu(X) = 1$. Several ways to automatically set the $2^n - 2$ remaining values [45] exist. The main problem is giving a value to sets having more than three elements while keeping the monotonic property of the integral and at the same time finding an approximation of the fuzzy measure that minimizes the error criterion. Generally, the problem is translated to another minimization
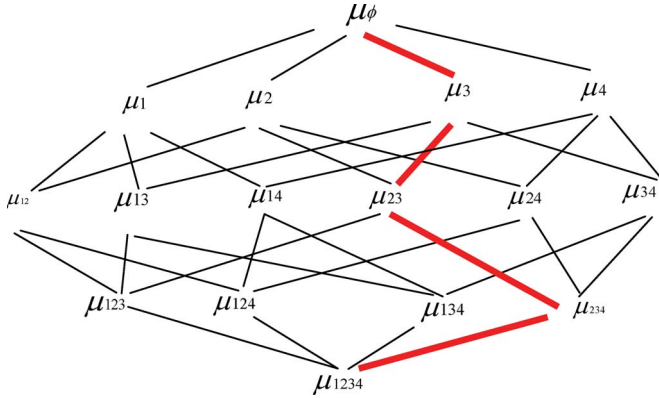
Fig. 4.   Example of a lattice associated to four DC.

problem which is usually solved using the Lemke method. Grabisch [16] has shown that such an approach may be inconsistent when using a low number of samples. In this instance (ill-conditioned matrices), the constraint matrix become parsed when the set of learning data grows, causing undesired behavior of the algorithm. To overcome these problems, an optimal approach based on a gradient algorithm with constraints, which is an extension of Muroshi and Sugeno's [46] method, has been proposed in [47]. It assumes that, in the absence of any information, the most reasonable way of aggregation is the arithmetic mean, i.e., the Choquet integral with respect to an additive equally distributed fuzzy measure. In this case, a lattice description is associated to the fuzzy measure under consideration. An example of the lattice representation associated with four DC is shown in Fig. 4.

This algorithm tries to minimize the mean square error between the values of the Choquet integral with respect to the fuzzy measure being learned and the expected values. For a training sample, the parameter vector is the current value of the fuzzy measure along the determined path by the ordering of the training vector coordinates.

See for instance the bold path in Fig. 4 corresponding to $\varphi(x_1) \leq \varphi(x_4) \leq \varphi(x_2) \leq \varphi(x_3)$.

This parameter vector is translated along the gradient direction with a magnitude proportional to the error, thus updating the values along the path. This means that coefficients of the fuzzy measure, which are not related to the data, are kept as near as possible to the equilibrium point. Thus, this algorithm is still efficient when training data set is limited. It also has low computing time and low memory cost. The algorithm, which is described in [47], can be implemented easily.

D. Indexes

Once the fuzzy measure is learned, it is possible to interpret the contribution of each DC in the final decision. Several indexes can be extracted from the fuzzy measure, helping to analyze the behavior of DC [16].

Importance Index: The importance of each DC, also called the Shapley index, is based on the definition proposed by Shapley in game theory [48] and is put back into fuzzy measure context by Murofushi and Soneda [16], [49]. Let a fuzzy

measure μ and a criterion i be considered. The main expression is given in (16) and so on for each DC

$$\sigma(\mu, i) = \frac{1}{n} \sum_{t=0}^{n-1} \frac{1}{\binom{n-1}{t}} \sum_{\substack{T \subseteq X \setminus i \\ |T| = t}} [\mu(T \cup i) - \mu(T)]. \quad (16)$$

The Shapley value can be interpreted as a weighted average value of the marginal contribution $\mu(T \cup i) - \mu(T)$ of criteria i alone in all combinations. A property worthy to be noted is that the sum of the indexes of all DC is equal to one; in other words, $\sum_{i=1,n} \sigma(\mu, i) = 1$. Hence, a DC with an importance index value less than $1/n$ can be interpreted as low impact in the final decision. Otherwise, an importance index greater than $1/n$ describes an attribute more important than the average.

Interaction Index: The interaction index, also called the Murofushi and Soneda index [16], [49], represents the positive or negative degree of interaction between two DC. If the fuzzy measure is nonadditive, then some sources interact. The marginal interaction between i and j, conditioned to the presence of the elements of combination $T \subseteq X \setminus ij$, is given by

$$(\Delta_{ij}\mu)(T) = \mu(T \cup ij) + \mu(T) - \mu(T \cup i) - \mu(T \cup j). \quad (17)$$

After averaging this criterion over all the subsets of $T \subseteq X \setminus ij$, the assessment of the interaction index of DC i and j is defined by (values in $[-1, 1]$)

$$I(\mu, ij) = \sum_{T \subseteq X \setminus ij} \frac{(n - t - 2)! t!}{(n - 1)!} (\Delta_{ij}\mu)(T). \quad (18)$$

This continues with any pair (i, j) with $i \neq j$. Obviously, the indexes are symmetrical, i.e., $I(\mu, ij) = I(\mu, ji)$.

A positive interaction index, for two DC i and j, means that the importance of one DC is reinforced by the second one. In other words, both DC are complementary, and their combined use makes the final decision better. The magnitude of this complement is given by the value of the index. A negative interaction index indicates that sources are antagonists and that their combined use impairs the final decision.

E. Automatic Extraction of Subsets of DC

Once the lattice is known, the individual performance of each DC is analyzed in the produced fuzzy measure. This analysis is performed using the importance and interaction indexes. The DC having the least influence in the final decision, and interacting the least with the other criteria, are assumed to blur the final decision.

A two-step selection scheme has been implemented to discard such DC. First, the Shapley value is scaled by the number n of DC. A DC with a scaled importance index greater than one describes a more important DC than the average.

The set of low significant criteria S_L, having an importance index lower than one, is selected

$$S_L = \{k/n \cdot \sigma(\mu, k) < 1\}. \quad (19)$$

Then, the subset of DC having the least positive synergy with the others is extracted from S_L. For each criterion S_{iL}, the values of its interaction with others are averaged to estimate its

global interaction. Finally, the subset of criteria to be removed MS_L is composed of the criteria from S_L, having an interaction index lower than the mean of the interaction indexes of all criteria of S_L

$$MS_L = \left\{ k / \sum_{j=1,n} I(\mu, k_j) < m \right\}, \qquad k \in S_L \qquad (20)$$

with the global mean interaction index

$$m = 1/|S_L| \Sigma k \epsilon S_L \Sigma_{j=1,n} I(\mu, k_j). \qquad (21)$$

IV. EXPERIMENTAL SETUP

A. Putting the FRC and the Feature Selection Scheme Together

The FRC and the suitable feature selection are embedded in a pattern recognition system. In such a system, a large set of features provides a large amount of fuzzy rules which are hardly exploitable. The selection process aims at decreasing the number of rules by discarding weak parameters while keeping the interesting recognition rates. First, the inference engine is run using initial features. From this first set of features and associated DC, a set of learning samples is determined.

Then, the fuzzy measure is obtained with respect to the Choquet integral. Indexes are extracted to determine the least representative DC. The recognition model is generated without the first least representative features and tested. The process is iterated while it remains less representative features. An overall algorithm of this fuzzy rule iterative feature selection (FRIFS) method is given below.

Algorithm of the FRIFS method:
Input:
  N: $\{1, \ldots, n\}$ classes
  P: $\{p_1, \ldots p_{|P|}\}$ parameters
  L_P: set of learning samples with $|p|$ parameters
  T_P: set of test samples with $|p|$ parameters
Begin
  - *Learn* $\mathrm{FRC}(N, Lp)$;
  - *Rate* $\leftarrow$ *apply* $\mathrm{FRC}(N, Tp)$;
  - *Stop* $\leftarrow$ *false*;
  While not *Stop* **do**
    - *Set* Capacity C to arithmetic sum considering $|p|$ criteria
    - *Calculate* alternatives from L_P
    - *Learn* C using Grabish' Algorithm
    - *Extract* Shapley value S_L
    - *Extract* weak parameters $M_{SL} = \{p'_1, \ldots, p'_{|MSL|}\}$
    - *Rate#* $\leftarrow Max_{i=1,|MSL|}\{\mathrm{FRC}(N, T_{P \setminus \{p'i\}})\}$
    If Rate# $>$ Rate **then**
      $P \leftarrow P \setminus \{P'_{ArgMaxi=1,|MSL|}\{\mathrm{FRC}(N, TP \setminus \{p'i\})\}\}$
      $Rate \leftarrow Rate\#$
    Else
      $Stop \leftarrow true$
    Endif
  Endwhile
  - Define set of final rules: $\mathrm{FRC}(N, Lp)$;
End

Methods	Data Sets	Error rates	F	Data Sets	Error rates	F
SVM	Balance	**21.12**	**4**	Diabetes	19.12	7
BFS		**21.12**	**4**		19.12	7
FFS		**21.12**	**4**		**18.46**	**6**
FRIFS		**21.12**	**4**		19.12	7
SVM	Iris	2.67	4	Glass	8.87	7
BFS		2.67	4		8.87	8
FFS		**2.67**	**3**		8.87	7
FRIFS		**2.67**	**3**		**8.87**	**6**
SVM	Liver	26.01	5			
BFS		26.01	5			
FFS		26.01	5			
FRIFS		**22.45**	**4**			

E = error rate. F = number of features. SVM = support vector machine
SBFS = backward feature selection. SFFS = forward feature selection
FRIFS = proposed method

B. Academic Validation of the FRIFS Method

To evaluate the effectiveness of this proposal, comparisons of five benchmarks from the UCI have been performed. These tests demonstrate the performance of the proposed feature selection method and the proposed classification method.

The five benchmarks are the following.

1) UCI Balance benchmark: It is composed of 625 samples divided into three classes. For each sample, four different features are calculated.
2) UCI Iris benchmark: It is composed of 150 samples divided into three classes. For each sample, four different features are calculated.
3) UCI Liver benchmark: It is composed of 345 samples divided into two classes. For each sample, six different features are calculated.
4) UCI Diabetes benchmark: It is composed of 768 samples divided into two classes. For each sample, eight different features are calculated.
5) UCI Glass benchmark: It is composed of 214 samples divided into six classes. For each sample, nine different features are calculated.

The error rates presented in Table I are obtained in memorization for the previous benchmarks, i.e., all samples are used to generate the recognition model, and the recognition rate is evaluated on all samples. The comparison done in Table I is made with four feature selection methods. The first, named FRIFS, implements to the FRIFS algorithm, the second is based on SVM [50], the third, based on a sequential backward feature selection, is named SBFS, and the fourth, based on a sequential forward floating selection, is named SFFS [51]. The last three methods are used to situate the efficiency of the proposed one. Indeed, SBFS and SFFS methods are well adapted to "small scale" context. They also give better results for UCI databases than others such as k-NN, Bayesian classifier, or the DTM [42]. For instance, the error rates of memorization obtained with DTM are 22.58, 7.98, 26.74, 20.03, and 9.65 for Balance, Iris, Liver, Diabetes, and Glass UCI databases, respectively. Other comparisons are shown in [52].

For these comparative tests, only the feature selection method is changed, not the classifier which remains the FRC.

In three benchmarks (Iris, Liver, and Glass), the number of features is less than the other tested feature selection algorithms (SVM, SBFS, and SFFS). In the case of the Balance benchmark, it can be noted that the feature has not been removed for all the methods. In the case of the Diabetes benchmark, FRIFS proposal is equivalent to the SVM and the SBFS algorithms, but the SFFS method is better. In this case, error rates considered, the difference is only due to five samples incorrectly classified out of 768 samples.

V. INDUSTRIAL APPLICATIONS

This paper is made in an industrial context. The proposed approach has been tested on three industrial data sets. These sets correspond to the components of the characteristic vector. The characteristic vector is calculated on the region of interest which has been determined to be defective. The segmentation step is performed by the industrial partner, and this paper aims essentially for defect identification. This industrial context makes the generation of the data sets more difficult. Indeed, for time constraints and availability, the labeling of each defect by an expert, a tiresome and long operation is inevitably reduced. Moreover, the defect number per class is not constant because it depends on the production at the time of the measure. Thus, the data sets are small and generally incomplete.

A. Presentation of the Industrial Data Sets

The three data sets, which have been tested, correspond to two different fabrics. The proposed approach has been elaborated on the first two data sets which relate to the same fabric, and the third has been used to check the proposed FRIFS method. For confidential reasons, the used features cannot be explicitly named. However, it can be noted that industrialist uses only attribute forms, sizes, and colors which have a signification, i.e., interpretable for a human who must tune the system.

The first data set, named Fibre1, contains six output classes (from C1 to C6) with 570 training samples. This set is decomposed in the following way: C1: 5, C2: 118, C3: 274, C4: 82, C5: 34, C6: 57. It can be noted that class C6 corresponds to invalid defects, i.e., detected defects which are not validated like defects from the customer. Thus, this class is strongly heterogeneous.

The second data set, named Fibre2, consists of a more extended version of the first data set. It contains the same six output classes with 618 training samples: C1: 12, C2: 188, C3: 230, C4: 131, C5: 2, C6: 55. In both data sets, classes C2, C3, and C4 allow consideration for a consistent training step, which is not the case for the other classes with, in the worst case, two or five representative samples.

The third set is decomposed in the following way: C1: 1, C2: 2, C3: 97, C4: 138, C5: 29, C6: 12, C7: 2, C8: 12, C9: 11, C10: 37, C11: 7, C12: 48. Class C12 corresponds to invalid defects.

TABLE II
COMPARISON BETWEEN THE DIFFERENT METHODS OF FEATURE SELECTION-GENERALIZATION RATES OBTAINED WITH FIBRE1 AND FIBRE2 DATA SETS

Data Sets		Fibre1				Fibre2			
Methods		SVM	SBFS	SFFS	**FRIFS**	SVM	SBFS	SFFS	**FRIFS**
9 features	rate	90.13				92.11			
8 features	without	P5	**P3**	**P3**	**P3**	P3	P3	P3	P3
	rate	89.48	**90.94**	**90.94**	**90.94**	92.63	92.63	92.63	92.63
7 features	without	P2	P0	P0	P0	P7	P5	P5	P5
	rate	86.89	90.29	90.29	90.29	92.46	92.63	92.63	92.63
6 features	without	P8	P5	P5	P5	P6	P8	P8	P8
	rate	88.67	89.48	89.48	89.48	90.00	92.63	92.63	92.63
5 features	without	P7	P8	P8	P8	P8	P2	P2	P0
	rate	81.88	86.73	86.73	86.73	83.51	90.00	90.00	92.81
4 features	without	P3	P7	P7	P7	P2	P8	P8	**P7**
	rate	81.23	70.07	70.07	70.07	81.05	81.05	81.05	**92.81**
3 features	without					P1	P1	P1	P2
	rate					74.74	74.74	74.74	90.88

SVM = support vector machine, SBFS = backward feature selection
SFFS = forward feature selection, FRIFS = Proposed Method

B. Characteristic Parameters

The constraints of the system require the use of simple features which can be quickly calculated. The counterpart of this simplicity is less of a discriminating aspect for some parameters or a redundancy for others. The number of features was initially set to 11. Local industrial expertise made it possible to delete two parameters which did not bring anything to the system and strongly decreased the results. The remaining parameters are equal to nine. The use of these nine parameters generates a consequent number of rules. In addition, some of them are strongly correlated with others.

C. Extraction of the Most Relevant Features

The four preceding feature selection methods have been applied to the data sets. The efficiency of the FRC has been evaluated by comparing the recognition rates obtained with the nine parameters. Table III gives these rates for SFFS, SBFS, SVM, and FRC (based on the FRIFS method). Moreover, other experiments have been made with the DTM [42]. The obtained rates are (in percent) 98.01, 96.11, and 90.71 for Fibre1, Fibre2, and Fabric industrial databases, respectively.

Table II summarizes the results that were obtained for the two data sets (Fibre1 and Fibre2). For these tests, the learning database consists of 33% of data sets. The remaining part (66% of data sets) is used to do the generalization step.

The columns of the Fibre1 data set present the recognition rates obtained by using this data set for the learning step and by applying the generated model to the Fibre2 data set, and inversely for the columns of the Fibre2 data set. Thus, these rates correspond to the generalization of the model. The proposed approach is iterative. It uses the recognition rate as an ending criterion. The value of this criterion has been empirically given to be 90%.

Memorization recognition rates are presented in Table III for Fibre1, Fibre2, and Fabric data sets.

The FRIFS method gives the best results and a better feature selection in comparison with the other methods. The other algorithms stop more quickly (except for the Fibre1 data set

Data Sets		Fibre1				Fibre2				Fabric			
Methods		SVM	SBFS	SFFS	**FRIFS**	SVM	SBFS	SFFS	**FRIFS**	SVM	SBFS	SFFS	**FRIFS**
9 features	Rate	98.95				97.74				90.96			
8 features	without	P5	P3	P3	P3	P3	P3	P3	P3	P1	P1	P1	P1
	Rate	98.42	98.77	98.77	98.77	97.57	97.57	97.57	97.57	90.45	90.45	90.45	90.45
7 features	without	P2	P0	P0	P0	P7	P5	P5	P5	P0	P0	P0	P0
	Rate	98.07	98.25	98.25	98.25	96.44	97.57	97.57	97.57	90.96	90.96	90.96	90.96
6 features	without	P8	P5	P5	P5	P6	P8	P8	P8	P4	P8	P8	**P5**
	Rate	97.72	97.90	97.90	97.90	90.29	97.41	97.41	97.41	85.18	87.19	87.19	**91.46**
5 features	without	P7	P8	P8	P8	P8	P2	P2	P0	P8	P4	P4	P8
	Rate	97.02	96.84	96.84	96.84	82.85	90.29	90.29	97.41	74.62	74.62	74.62	84.67
4 features	without	P3	P7	P7	P7	P2	P8	P8	P7				
	Rate	96.97	96.14	96.14	96.14	77.02	77.02	77.02	96.93				
3 features	without					P1	P1	P1	P2				
	Rate					76.38	76.38	76.38	92.88				

SVM = support vector machine, SBFS = backward feature selection
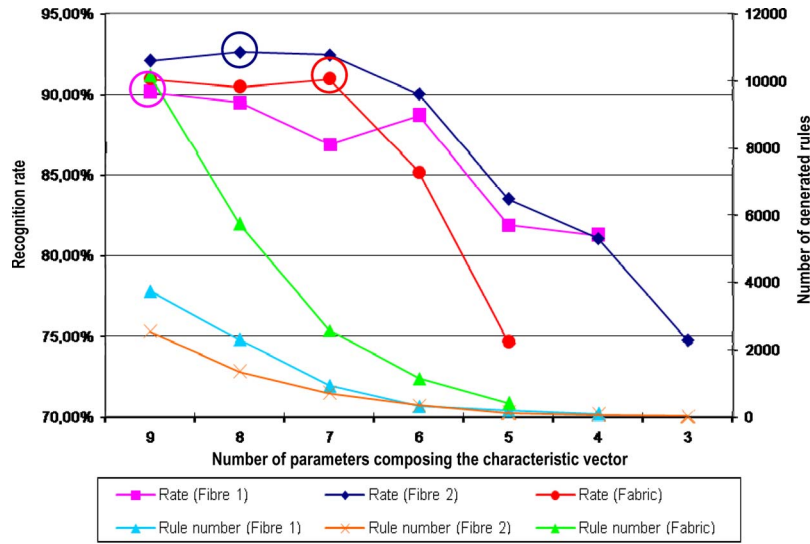SFFS = forward feature selection, FRIFS = Proposed Method

Fig. 5. Evolution of the recognition rates and the rule number for the SVM method.

with the BFS and FFS methods) than the proposed approach by providing a less satisfactory parameter set.

It can be noted that the suppression order of the nonrelevant parameters is different, not only among the four methods but also according to the data set chosen for the training step. On the other hand, for the FRIFS approach, the same four parameters are selected (even five, if an ending criterion is chosen lower than 90%). It is not the case for the other methods.

Thus, the FRIFS method seems to be more efficient and provides a more stable characteristic vector composed of more significant features.

Fig. 7 shows the stability of the recognition rate and its progressive diminishing, which is neither the case for the SVM (Fig. 5) nor for the SBFS nor SFFS methods (Fig. 6). The results obtained with the Fibre1 and Fibre2 data sets have been validated on the Fabric data set (Table II). In this case, the proposed method eliminates one more parameter. This also implies a decrease in the rule set complexity while improving the recognition rate.

In connection with the complexity, the values shown in Figs. 5–7 correspond to the numbers of active rules, i.e., the number of rules with the CF confident coefficient (see Section II-C) that is strictly higher than zero. These graphs show a decrease in the active rule number proportionally to the global rule number (functions of the size of the characteristic vector and the fuzzification term number chosen for each parameter). However, on all of these data sets, the active rule number is higher for the proposed FRIFS method, which seems to mean that the selected parameters are more relevant and that the obtained model is more accurate.

Finally, it can be noted that 48 active rules are obtained (per 80 rules) for the Fibre2 data set. Thus, the system becomes interpretable.

VI. CONCLUSION AND FUTURE WORKS

The FRIFS method proposed in this paper is based on the analysis of a training data set in three steps. The first step, representing the initialization of the method, allows for the choice of the first subset of parameters starting from an analysis of the data typicality. The second and third steps are the iterative parts of the method and reduce the dimension problem while keeping a high recognition rate.

The FRIFS approach seems to be more robust because it keeps homogeneity and better stability according to the recognition rates and to the number of selection rules.
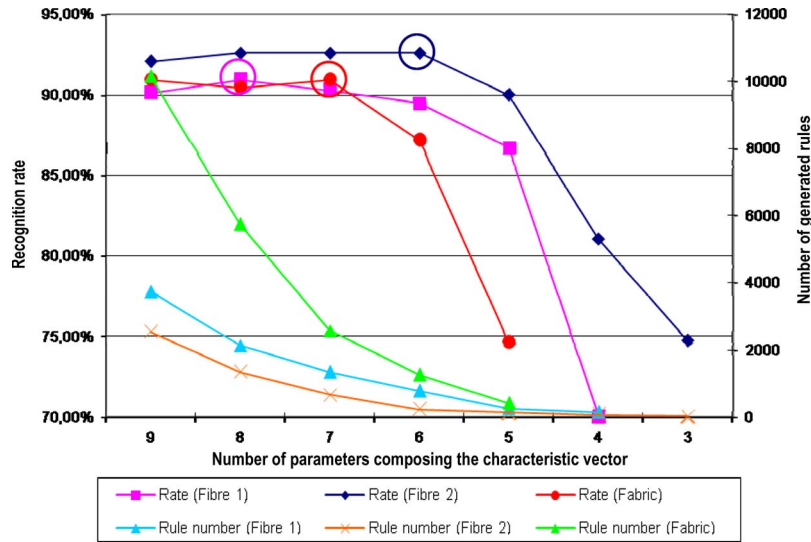
Fig. 6.   Evolution of the recognition rates and the rule number for the SBFS and SFFS methods.
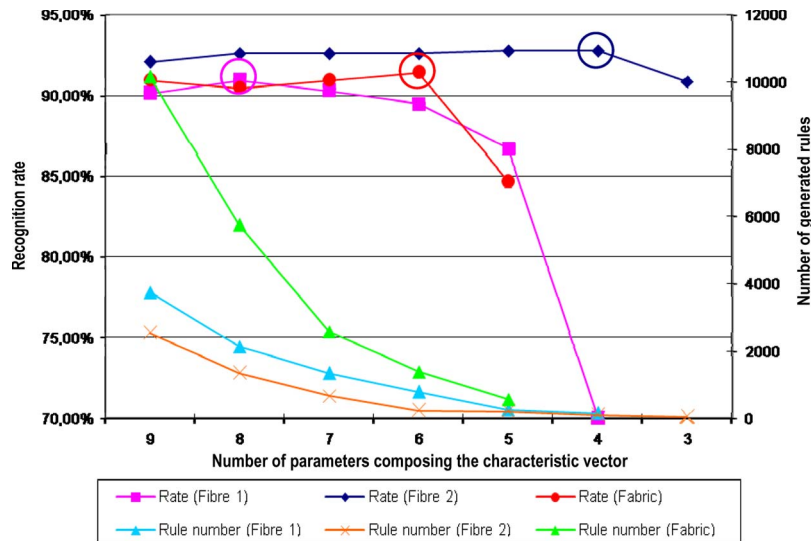

Fig. 7.   Evolution of the recognition rates and the rule number for the FRIFS method.

The experimental results have shown that the proposed method allows the choice of an optimal subset of parameters, increasing the recognition rate in comparison with the choice carried out by expertise and keeping a certain degree of interpretability in the model.

Several experimental results, made on UCI databases, allow for comparison on well-known databases and show the efficiency of the proposed method on different data sets. Moreover, the FRIFS method has been successfully applied in another industrial domain for recognizing wood defects [53].

However, with 257 and 201 active rules (out of the 960 generated) if the removal of the same four parameters on Fibre1 and Fibre2 data sets are considered, the model is not as easily interpretable.

Using a multicriteria model should not exceed around ten criteria [54]. Due to the setting of the algorithm to arithmetic mean, the gradient algorithm is still usable, handling around 20 criteria. Nonetheless, uncertainty about the interpretability of the model remains, considering the large size of the associated lattice. Currently, two-additive Choquet integral schemes are under consideration to increase the scalability of the approach.

Thus, further investigations aim to reduce the number of generated rules. An extension of the proposed FRIFS method aims to analyze each class and not all the training data set, as currently carried out.

Other possible investigations consist in coupling the fuzzy decision tree (FDT) method with the FRC. In this case, it is envisaged to use a tree version of the FRC. This model is based on a tree structure of fuzzy inference systems where each configuration is led by expertise and integrated expert knowledge [55]. Its main advantage is to reduce the number of rules in order to increase the interpretability of the models. Moreover, the rule set given by Ishibuchi's algorithm should be utilized according to their CF confident degree and to merge them with the rule set given by an FDT method. This improvement is the basis of an adapted feature selection to use a higher level decision-making process which gives better results in terms of recognition rate and interpretability.

Acknowledgment

The authors would like to thank their industrial partner who provided the data sets and knowledge concerning the framework.

References

[1] M. Kudo and J. Sklansky, "Comparison of algorithms that select features for pattern classifiers," *Pattern Recognit.*, vol. 33, no. 1, pp. 25–41, Jan. 2000.

[2] H. Zhang and G. Sun, "Feature selection using tabu search method," *Pattern Recognit.*, vol. 35, no. 3, pp. 701–711, Mar. 2002.

[3] O. Melnik, Y. Vardi, and C. H. Zhang, "Mixed group ranks: Preference and confidence in classifier combination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 973–981, Aug. 2004.

[4] B. Littewood and D. Miller, "Conceptual modeling of coincident failures in multiversion software," *IEEE Trans. Softw. Eng.*, vol. 15, no. 12, pp. 1596–1614, Dec. 1989.

[5] T. K. Ho, "Multiple classifier combination: Lessons and next steps," in *Hybrid Methods in Pattern Recognition*, A. Kandel and H. Bunke, Eds. Singapore: World Scientific, 2002.

[6] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, Aug. 1996.

[7] R. E. Schapire, Y. Fruend, P. Bartlett, and W. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods," *Ann. Stat.*, vol. 26, no. 5, pp. 1651–1689, Oct. 1998.

[8] S. T. Hadjitodorov, L. I. Kuncheva, and L. P. Todorova, "Moderate diversity for better cluster ensembles," *Inf. Fusion*, vol. 7, no. 3, pp. 264–275, Sep. 2006.

[9] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2001.

[10] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[11] D. Ruta and B. Gabrys, "An overview of classifier fusion methods," *Comput. Inf. Syst.*, vol. 7, no. 1, pp. 1–10, 2000.

[12] Z. Stejic, Y. Takama, and K. Hirota, "Mathematical aggregation operators in image retrieval: Effect on retrieval performance and role in relevance feedback," *Signal Process.*, vol. 85, no. 2, pp. 297–324, Feb. 2005.

[13] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.

[14] R. Yager, "On ordered weighted averaging aggregation operators in multi-criteria decision making," *IEEE Trans. Syst., Man, Cybern.*, vol. 18, no. 1, pp. 183–190, Jan./Feb. 1988.

[15] J.-L. Marichal, "Aggregation of interacting criteria by means of the discrete Choquet integral," in *Aggregation Operators: New Trends and Applications*. Heidelberg, Germany: Physica-Verlag GmbH, 2002, pp. 224–244.

[16] M. Grabisch, "The application of fuzzy integrals in multicriteria decision making," *Eur. J. Oper. Res.*, vol. 89, no. 3, pp. 445–456, Mar. 1996.

[17] S. B. Cho and J. H. Kim, "Combining multiple neural networks by fuzzy integral for robust classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 2, pp. 380–384, Feb. 1995.

[18] H. Tahani and J. M. Keller, "Information fusion in computer vision using the fuzzy integral," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, no. 3, pp. 733–741, May/Jun. 1990.

[19] Y. S. Choi, D. Kim, and R. Krishnapuram, "Relevance feedback for content-based image retrieval using the Choquet integral," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2000, vol. 2, pp. 1207–1210.

[20] S. Chang and S. Greenberg, "Application of fuzzy-integration-based multiple-information aggregation in automatic speech recognition," in *Proc. IEEE Conf. Fuzzy Integr. Process.*, Beijing, China, 2003. CD-ROM, (8 p.).

[21] E. Schmitt, V. Bombardier, and P. Charpentier, "Self-fuzzification method according to typicality correlation for classification on tiny data sets," in *Proc. 16th Int. Conf. FUZZ-IEEE*, London, U.K., 2007, pp. 1072–1077.

[22] J. Rendek and L. Wendling, "Extraction of consistent subsets of descriptors using Choquet integral," in *Proc. 18th Int. Conf. Pattern Recog.*, Hong Kong, China, 2006, vol. 3, pp. 208–211.

[23] C. J. Mertz and P. M. Murphy, *UCI Repository of Machine Learning Databases*. Irvine: Univ. California, 1998. [Online]. Available: http://www.ics.uci.edu.pl/mlearn/MLRespository.html

[24] E. Schmitt, C. Mazaud, V. Bombardier, and P. Lhoste, "A fuzzy reasoning classification method for pattern recognition," in *Proc. 15th Int. Conf. FUZZ-IEEE*, Vancouver, BC, Canada, 2006, pp. 1078–1085.

[25] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[26] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Inf. Sci.*, vol. 8, no. 3, pp. 199–249, 1975.

[27] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—II," *Inf. Sci.*, vol. 8, no. 4, pp. 301–357, 1975.

[28] O. Cordon, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, "Ten years of genetic fuzzy systems: Current framework and new trends," *Fuzzy Sets Syst.*, vol. 141, no. 1, pp. 5–31, Jan. 2004.

[29] T. Kempowsky, A. Subias, and J. Aguilar-Martin, "Process situation assessment: From a fuzzy partition to a finite state machine," *Eng. Appl. Artif. Intell.*, vol. 19, no. 5, pp. 461–477, Aug. 2006.

[30] F. A. T. De Carvalho, "Fuzzy c-means clustering methods for symbolic interval data," *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 423–437, Mar. 2007.

[31] J. Forest, M. Rifqi, and B. Bouchon-Meunier, "Class segmentation to improve fuzzy prototype construction: Visualization and characterization of non homogeneous classes," in *Proc. IEEE World Congr. Comput. Intell.*, Vancouver, BC, Canada, 2006, pp. 555–559.

[32] D. Dubois and H. Prade, "Fuzzy rules in knowledge-based systems—Modelling gradedness, uncertainty and preference," in *An Introduction to Fuzzy Logic Application in Intelligent Systems*. Dordrecht, The Netherlands: Kluwer, 1992, pp. 45–68.

[33] D. Dubois and H. Prade, "What are fuzzy rules and how to use them?" *Fuzzy Sets Syst.*, vol. 84, no. 2, pp. 169–185, Sep. 1996.

[34] D. Dubois, H. Prade, and L. Ughetto, "Checking the coherence and redundancy of fuzzy knowledge bases," *IEEE Trans. Fuzzy Syst.*, vol. 5, no. 3, pp. 398–417, Aug. 1997.

[35] J. M. Mendel, "Fuzzy logic systems for engineering: A tutorial," *Proc. IEEE*, vol. 83, no. 3, pp. 345–377, Mar. 1995.

[36] M. Sugeno, "An introductory survey of fuzzy control," *Inf. Sci.*, vol. 36, no. 1/2, pp. 59–83, Jul./Aug. 1985.

[37] H. Ishibuchi, K. Nozaki, and H. Tanaka, "Distributed representation of fuzzy rules and its application to pattern classification," *Fuzzy Sets Syst.*, vol. 52, no. 1, pp. 21–32, Nov. 1992.

[38] R. Alcala, J. Alcala-Fdez, F. Herrera, and J. Otero, "Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation," *Int. J. Approx. Reason.*, vol. 44, no. 1, pp. 45–64, Jan. 2007.

[39] O. Cordon, M. J. del Jesus, and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems," *Int. J. Approx. Reason.*, vol. 20, no. 1, pp. 21–45, Jan. 1999.

[40] M. R. Berthold, "Mixed fuzzy rule formation," *Int. J. Approx. Reason.*, vol. 32, no. 2, pp. 67–84, Feb. 2003.

[41] C. Marsala, "Fuzzy decision trees to help flexible querying," *Kybernetika*, vol. 36, no. 6, pp. 689–705, 2000.

[42] D. Michie, D. J. Spiegelhalter, and C. Taylor, Eds., *Machine Learning, Neural and Statistical Classification*. New York: Ellis Horwood, 1994.

[43] H. Ishibuchi, K. Nozaki, and H. Tanaka, "A simple but powerful heuristic method for generating fuzzy rules from numerical data," *Fuzzy Sets Syst.*, vol. 86, no. 3, pp. 251–270, Mar. 1997.

[44] H. J. Zimmermann, *Fuzzy Set Theory and Its Applications*, 4th ed. Norwell, MA: Kluwer, 2001.

[45] M. Grabisch and J. M. Nicolas, "Classification by fuzzy integral—Performance and tests," *Fuzzy Sets Syst.*, vol. 65, no. 2/3, pp. 255–271, Aug. 1994.

[46] T. Murofushi and M. Sugeno, "A theory of fuzzy measures: Representations, the Choquet integral, and null sets," *J. Math. Anal. Appl.*, vol. 159, no. 2, pp. 532–549, 1991.

[47] M. Grabisch, "A new algorithm for identifying fuzzy measures and its application to pattern recognition," in *Proc. 7th IEEE Int. Joint Conf. Fuzzy Syst.*, Yokohama, Japan, Mar. 1995, pp. 145–150.

[48] L. Shapley, "A value for n-person games," in *Contributions to the Theory of Games, Annals of Mathematics Studies*, H. Khun and A. Tucker, Eds. Princeton, NJ: Princeton Univ. Press, 1953, pp. 307–317.

[49] T. Murofushi and S. Soneda, "Techniques for reading fuzzy measures(iii): Interaction index," in *Proc. 9th Fuzzy Syst. Symp.*, Sapporo, Japan, May 1993, pp. 693–696. (In Japanese).

[50] Y. Grandvalet and S. Canu, "Adaptive scaling for feature selection in SVMs," in *Neural Information Processing System*. Cambridge, MA: MIT Press, 2003.

[51] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, Nov. 1994.

[52] M. Atif Tahir, A. Bouridane, and F. Kurugollu, "Simultaneous feature selection and feature weighting using Hybrid Tabu Search/K-nearest

neighbor classifier," *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 438–446, Mar. 2007.

[53] C. Mazaud, J. Rendek, V. Bombardier, and L. Wendling, "A feature selection method based on Choquet integral and typicality analysis," in *Proc. 16th Int. Conf. FUZZ-IEEE*, London, U.K., 2007, pp. 1073–1708.

[54] M. Grabisch, J. Duchêne, F. Lino, and P. Perny, "Subjective evaluation of discomfort in sitting position," *Fuzzy Optim. Decision Making*, vol. 1, no. 3, pp. 287–312, 2002.

[55] V. Bombardier, C. Mazaud, P. Lhoste, and R. Vogrig, "Contribution of fuzzy reasoning method to knowledge integration in a defect recognition system," *Comput. Ind.*, vol. 58, no. 4, pp. 355–366, May 2007.

Vincent Bombardier was born in Nancy, France, in 1964. He received the Ph.D. degree in computer science from the University Henri Poincaré (UHP), Nancy, in 1991.

He is an Associate Professor with the École Supérieure d'Informatique et Applications de Lorraine, UHP Nancy 1 and a Researcher with the Centre de Recherche en Automatique de Nancy in the project team SYstèMes de Production Ambiants. Since 1992, its research works combine image processing (segmentation, color, texture, pattern recognition, etc.) and fuzzy set theory application (fuzzy logic, fuzzy rules, etc.). His studies mainly concern knowledge integration in pattern recognition problems, by merging qualitative (human skill) and quantitative information (sensor) in order to improve industrial automatic vision systems.

Emmanuel Schmitt was born in Schiltigheim, France, in 1980. He received the Engineering degree in telecommunications from the Institut des Sciences de l'Ingénieur de Toulon et du Var, La Valette du Var Cedex, France, in 2003, and the Ph.D. degree in automation, computer science, and signal processing from Université Henri Poincaré Nancy 1, Vandoeuvre-lès-Nancy Cedex, France, in 2007.

Since 2007, he has been a Development Engineer in automation. He continues to work with the Centre de Recherche en Automatique de Nancy, CNRS UMR 7039, Université Henri Poincaré—Campus Scientifique, Vandoeuvre-lès-Nancy Cedex on different fields like fuzzy logic, simplification of settings in pattern recognition, and image processing.

Laurent Wendling was born in Lunéville, France, in 1969. He received the Ph.D. degree in computer science from the University Paul Sabatier, Toulouse, France, in 1997. He received the "HDR" degree in 2006.

From 1993 to 1999, he was with the Toulouse Institute of Research in Computing Science in the field of pattern recognition. His current research topics are symbol recognition, feature selection, and relevance feedback techniques. He is currently an Assistant Professor with the École Supérieure d'Informatique et Applications de Lorraine, Université Henri Poincaré, Vandoeuvre-lès-Nancy, France. He is also a member of the Querying Graphics through Analysis and Recognition team, Laboratoire Lorrain de Recherche en Informatique et ses Applications in Nancy.