Relations among wavelet coefficients and features for ATR

Mieczyslaw M. Kokar
Department of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115
kokar@coe.neu.edu

Marek K. Malczewski
Department of Mechanical, Manufacturing and Industrial Engineering
Northeastern University
Boston, MA 02115
marekmal@coe.neu.edu

ABSTRACT

Wavelets have been used successfully for signal compression. A signal can be represented very concisely and with a high fidelity, by a set of wavelet coefficients. This suggests that wavelet coefficients can efficiently represent the contents of a signal and, consequently, could be used as features. Such features then can be used for signal classification. The quality of classification depends on the choice of the features. Fixing the set of features in both time and frequency domains results in the lack of invariance of the classification method with respect to translations and scaling of signals. In this paper we propose an approach that addresses this problem. We achieve this goal by using the following two techniques. First, our classification method test whether a specific relation among wavelet coefficients is satisfied by a given signal. And second, our method selects features dynamically, i.e., it searches for features that satisfy the relation. The relations are learned from a database of pre-classified signals. In this paper we provide the description of the relation learning approach and results of testing the approach on a simple scenario. The results of our simulations showed that this approach gives a higher classification accuracy than a similar approach based on a fixed set of features.

1. INTRODUCTION

Wavelets have been used successfully for signal compression. In other words, a signal can be represented very concisely and with a high fidelity, by a set of wavelet coefficients. This suggests that wavelet coefficients can efficiently represent the contents of a signal and, consequently, could be used as features. Such features then can be used for signal classification and recognition. This approach has been investigated by many authors. The problem is, though, which of the wavelet coefficients should be chosen as features? Some of the authors (e.g., Saito[7]) used an entropy based measure to select wavelet coefficients that are most useful as features. This is a fully data-driven approach in which a database of classes of signals is used to select the coefficients that, on the average, have the highest discrimination power. Korona and Kokar[5,4] used a theory-driven approach, i.e., they assumed that the features were selected by a human expert. Although they have shown that the knowledge-driven approach results in better features, the approach had to rely on the human-in-the-loop. Also, the features generated by both approaches were fixed in both time and frequency domains, i.e., once selected, the same coefficients were used later in the recognition process. In other words, the recognition based upon such features lacks the property of invariance with respect to both time and frequency. This makes both methods inappropriate for scenarios in which signals are that need to be classified as members of the same class are shifted in the time domain or are shifted in the frequency domain.

The main idea of the solution proposed in this paper consists of two parts. One, instead of selecting a fixed set of wavelet coefficients for the whole set of signals, select a most discriminant relation among a number of wavelet coefficients. Two, dynamically attempt to select a set of coefficients such that satisfy a relation for a given class of signals. If such a set of coefficients can be chosen, then the decision is made that the given signal is a member of that class. Otherwise, the decision is negative. A relation is defined as a k-ary relation among wavelet coefficients expressed as a set of k binary relations, where each such binary relation is either "equal", "less than" or "greater than".

The method described in this paper can be used for deriving classification rules for various kinds of signals. In particular, one could use the same approach to generate classification relations for two databases of signals coming from two different sensors. The derived relations could be either used for generating decisions, and then the decisions would need to be fused,[2,8] or the relations could be fused resulting in a single (fused) relation, which then could be used for classification of signals from two different sensors.[4]

In the paper we compare the quality of classification by this algorithm with the classification in which a fixed set of the Most Discriminant Wavelet Coefficients are used. Towards this goal, we generate two databases of signals, a training database and a testing database. We show that the misclassfication rate of our algorithm is significantly higher than for the Most Discriminant Wavelets Coefficients.

2. FORMULATION OF THE PROBLEM

The quality of feature-based recognition strongly depends on the quality of features used. However, there is a trade off between the generality of the algorithm and the quality of recognition. One can always select features that are best for a very narrow type of signals. A "good" set of features, however, is such that is applicable to a wide class of signals, rather than to a very narrow domain.

To formalize the recognition problem we assume the following scenario. First, there is a class of signals. In this paper we consider only 1D signals $s : T \rightarrow V$, where T is time and V is the value set of the signal. We consider features to be collections of pairs

$$\{< t_1, f(t_1) >, \ldots < t_n, f(t_n) >\} \tag{1}$$

where $t_1, \ldots t_n$ are values of the signal domain for a given signal s and $f(t_1), \ldots, f(t_n)$ are values of the feature function for the same signal s. We assume that the signals can be classified into classes, say $C_1, \ldots, C_K$ by *classification rules*. Often classification rules are represented as relational expressions, e.g.,

$$S_k(< t_1, f(t_1) >, \ldots < t_n, f(t_n) >) \Leftrightarrow f(t_1) \geq \tau_1 \wedge \ldots \wedge f(t_n) \geq \tau_n \tag{2}$$

where S_k $(k = 1, \ldots, K)$ is the classification function for the $k - th$ class and $\tau_1, \ldots, \tau_n$ are thresholds.

In this paper we focus on the selection of features $< t_1, f(t_1) >, \ldots < t_n, f(t_n) >$. In the simplest case these features can be fixed in the time domain, i.e., the values $t_1, \ldots t_n$ are fixed and used by the classification rules. In other words, the values of features are measured (or computed) for the same time instances of the signal's time coordinate. As we mentioned earlier, this is a rather strong constraint, meaning that such a classification rule is not invariant with respect to translations or scaling of the signal. In this paper we investigate a possibility of relaxing such a constraint. Towards this aim, we rewrite the classification rule of Equation 2 by replacing the fixed values $t_1, \ldots, t_n$ with the existential quantifier

$$\exists_{t_1,\ldots,t_n} S_k(< t_1, f(t_1) >, \ldots < t_n, f(t_n) >) \Leftrightarrow f(t_1) \geq \tau_1 \wedge \ldots \wedge f(t_n) \geq \tau_n \tag{3}$$

This rule is less restrictive than the rule given by Equation 2 in the sense that it can be satisfied whenever there exist the values of $t_1, \ldots, t_n$ such that satisfy the relational expression. The rule of Equation 2 on the other hand, requires that this expression needs to be satisfied for the fixed set of values $t_1, \ldots, t_n$. This means that some signals that would not satisfy Equation 2, may satisfy Equation 3. On the one hand, this may have a negative effect on the quality of classification (since it is easier now to satisfy such a rule). It might be so flexible that all the signals could satisfy a rule and thus the rule is not discriminative enough. But on the other hand, it also gives more flexibility, since this kind of rule can be used for classifying signals that are not exactly the same; signals that have some degree of "similarity" may be assigned to the same class. In this sense, this kind of classification rules can be invariant with respect to both translations and scalings of signals. To achieve the high degree of discrimination power we need to search for appropriate relational expressions; the conjunctive form like in Equation 2 might not be sufficient.

This formulation of the classification problem makes it more difficult to formulate classification rules. In the former formulation a classification rule could be "hand-crafted" by selecting a set of thresholds $\tau_1, \ldots, \tau_n$. Various tuning algorithms could be used for this purpose. In the formulation given by Equation 3, one needs to find the right hand side of the rule S_k that would replace the simple conjunctive form $f(t_1) \geq \tau_1 \wedge \ldots \wedge f(t_n) \geq \tau_n$. In this paper we present our initial experiments with this kind of problem. In particular, we show experiments with deriving classification rules through training.

3. CLASSIFICATION RELATIONS

In this section we describe the type of relational expressions that were used for classification. For finding classification relations we use a Relation Search (RS) algorithm. The relation is selected based upon a training database of signals. In this database, for each signal the signal's classification is given explicitly. The relation selected by the RS algorithm is called Most Discriminant Relation (MDR).

The number of possible relations for a given database of signals is huge. In fact, if we do not fix the arity of the relation, then it is even infinite. In our experiments we limited ourselves to a subset of all possible relations. First, we considered only finite relations of arity $n = 5$ (5-ary relations). More specifically, we focused on sets of five binary relations among wavelet coefficients, where each binary relation was one of the following:

$$B = \{<, =, >\} \tag{4}$$

An example of such a 5-ary relation Φ would be

$$\Phi = \{<, <, >, =, =\}. \tag{5}$$

We say that an ordered set of features $< t_1, f(t_1) >, \ldots < t_n, f(t_n) >$, where $t_1 < t_2 < t_3 < t_4 < t_5$, satisfies $\Phi = \{r_1, r_2, r_3, r_4, r_5\}$, where $r_1, \ldots, r_5 \in B$, if and only if there exist such values of $t_1, \ldots, t_5$ that $f(t_1) r_1 f(t_2)$ and $f(t_2) r_2 f(t_3)$ and $f(t_3) r_3 f(t_4)$ and $f(t_4) r_4 f(t_5)$ and $f(t_5) r_5 f(t_1)$. It is easy to calculate that there are $3^5 = 243$ possible combinations of such binary relations, i.e., there are 243 possible types of 5-ary relations like this.

The RS algorithm extracts the relation Φ given a set of features for a pre-classified set of signals. The relational formula described above can be used for classifying signals into two classes - those that satisfy the relation and those that don't. For each class of signals such a relation can be constructed giving a method for multi-class classification.

3.1. Entropy based classification

The relations described above don't have to be, and usually are not, disjoint. This means that a signal can satisfy multiple relational formulas. In the example considered above, the worst case would be if a signal satisfied all of the 243 formulas. We are interested in such formulas that are most discriminant, i.e., such that give most sharp classifications. Intuitively this means such classifications that have the least of the overlap among the various classes. To measure the quality of a classification rule we use the main idea of the entropy based classification, as for instance presented in a paper by Quinlan.[6]

Consider a two-class classification problem, i.e., we have two classes of signals, C_1 and C_2. Let p denote the number of signals in class C_1 and let n denote the number of signals in class C_2. The information associated with this classification is given by Quinlan[6]:

$$I(n,p) = -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\frac{n}{p+n} \tag{6}$$

Now consider a (less than perfect) classification rule Φ that splits the set of signals into two classes. Denote p_1 as the number of signals correctly classified as C_1, n_1 as the number of signals from C_2, but classified as C_1, p_2 as the number of signals correctly classified as C_2, and n_2 as the number of signals from C_1 (incorrectly) classified as C_2. Similarly as in the Quinlan's paper,[6] we can calculate the expected information of the classification

$$E(\Phi) = \sum_{i=1}^{2} \frac{p_i + n_i}{p+n} I(n_i, p_i) \tag{7}$$

and then the expected information gain due to this classification as

$$gain(\Phi) = I(p,n) - E(\Phi) \tag{8}$$

Our relation selection algorithm uses these measures for selecting best relations. It compares information gains for various relations and then selects the one with the highest information gain as the most discriminant relation.

4. EXPERIMENTAL SETUP

To address the problem described above we performed the experiments as shown in Figure 1. First of all, we prepared two databases of signals, a learning database and a testing database. The signals in the two databases have been transformed into the wavelet domain by applying Discrete Wavelet Packet Decomposition (DWPD).[1,3] In the next step, the Most Discriminant Bases (MDBs) were selected, one for the learning database and one for the training database, using the Best Discriminant Basis Algorithm (BDBA) as described in the previously referred papers.[1,7] After that, all the signals were represented in the selected MDBs. At this point we had as many coefficients as samples in each signal. The next step is to select a relation that would be used for classification. Since the goal was to assess our proposed approach against an approach in which a fixed set of features is used, we had to pursue two paths. In one path (the left most branch in Figure 1) we applied our relation selection algorithm to all of the coefficients of the MDB (for all signals). In the other path, we selected only five most discriminant elements of the MDB and used the coefficients of these MDB elements in the search for the most discriminant relation. The reason for this was to find out what is the gain in the recognition accuracy that can be attributed to the use of the relational formulation given by Equation 3 with respect to the formulation given by Equation 2. The result of the two left branches shown in Figure 1 were two relations. In the next step, these two relations were applied to classify the test database.

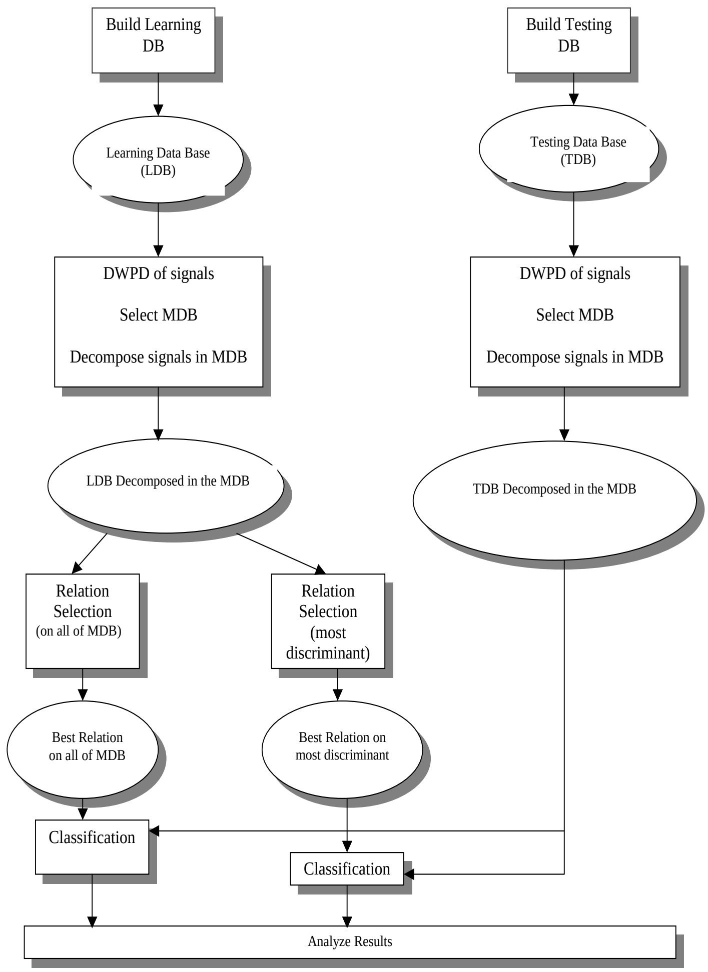
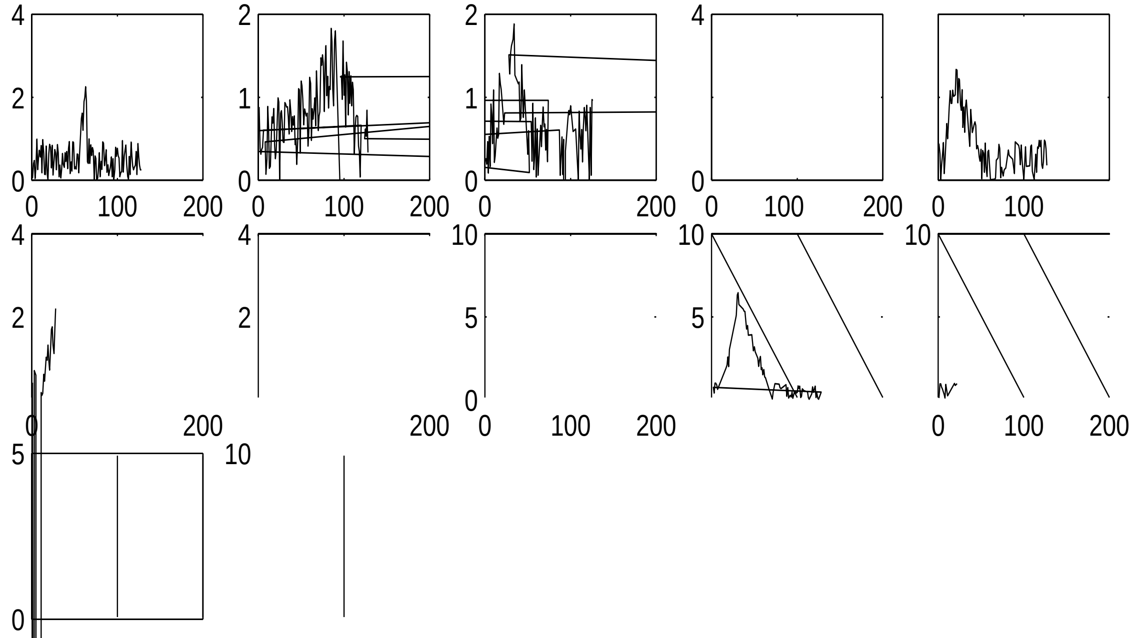


Figure 1. Experiments

5. SIMULATIONS

To obtain an initial assessment of the usability of this approach we developed two databases of signals, a training database and a testing database (see Figure 1). We generated 100 signals for the learning database and 600 signals for the training database. The signals were of two different classes, rectangular and triangular. In each of the databases there were 50% of signals from each of the two classes. Examples of triangular signals from the learning database are shown in Figure 2. As can be seen from this figure, the shapes of the signals were varied. The length of the triangle base, the location of the base in the time coordinate, the location of the apex of the triangle, both the height and the horizontal location with respect to the base, were all varied randomly. Similarly, for the rectangular signals, the size of the rectangle base, its location in the time coordinate, as well as the height, were varied randomly.



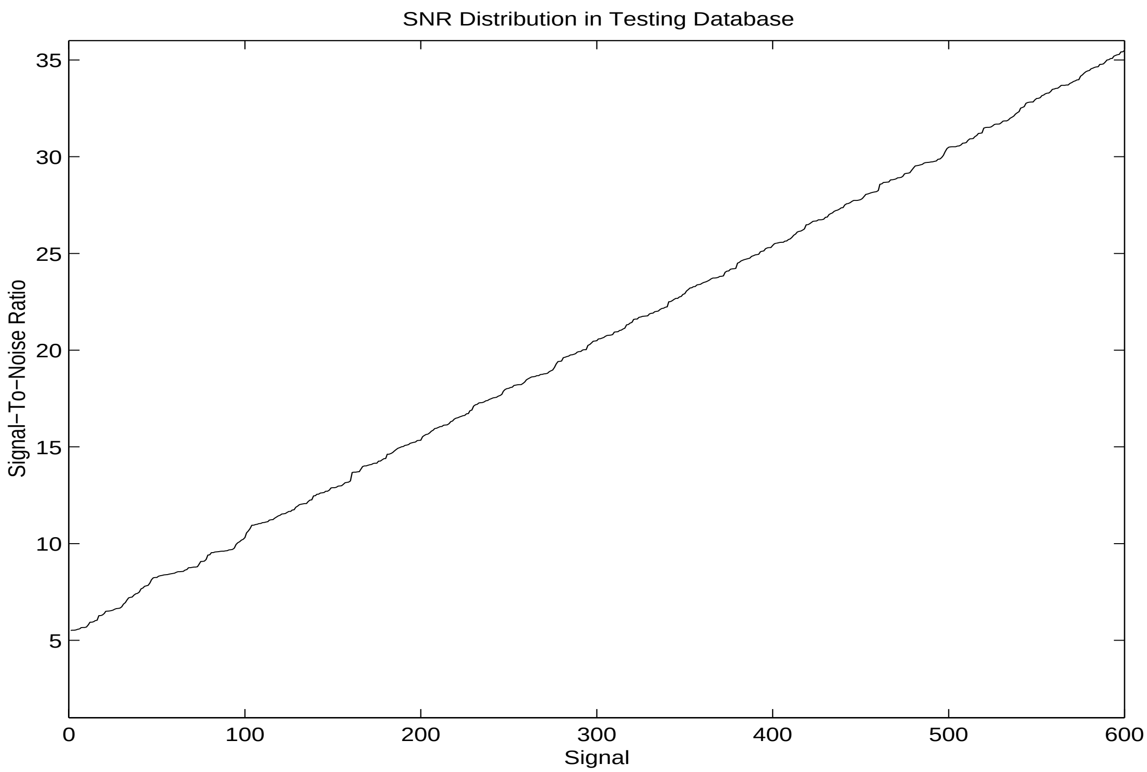


Figure 3. Distribution of SNR for Signals in the Learning Database

test database.

$$\delta = \frac{P}{P+N} \cdot 100 \tag{9}$$

We have run many simulation experiments, each addressed at a specific research question related to the method proposed in this paper. As we said before, our main goal was to assess the usability of the approach proposed in this paper. Towards this aim, we compared the classification accuracy of this method against the accuracy of recognition of a system in which the relations were learned, similarly as in the proposed system, but where the set of features was fixed. As shown in the middle branch of Figure 1, we apply the Best Discriminant Basis Algorithm (BDBA)[1,7] to select a collection of features that are best with respect to the entropy based measure used by the BDBA. Then we select the best relation using our approach. Finally, we classify the signals using both relations, one selected by our approach, and the other that used our approach to the set of features selected by the BDBA. The results of the classification (the classification accuracy) are plotted in Figure 4. The dashed line shows the classification accuracy of our method (using the MDR, i.e., most discriminant relations, it is marked as RS Relation in the figure) and the dashed-and-dotted line shows the classification accuracy of the relation among only the most discriminant features selected by the BDBA (in the figure it is called MDB Relation).

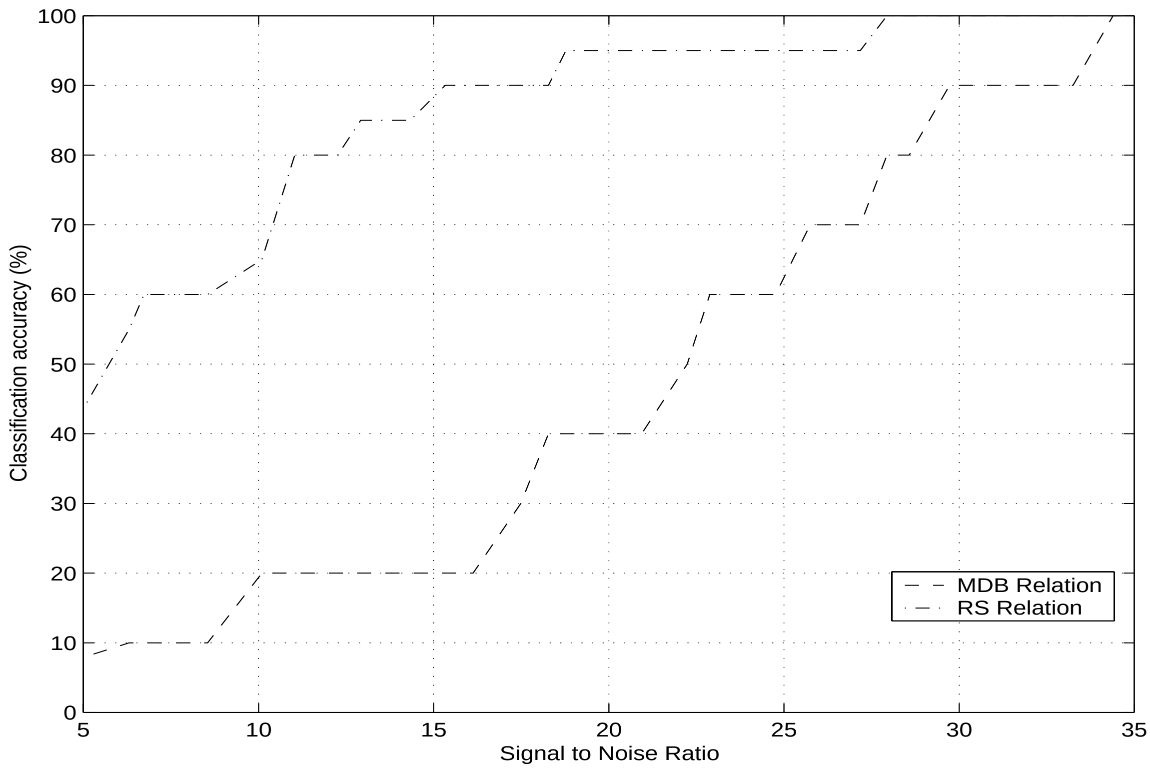


Figure 4. Classification Accuracy for Various Levels of Noise (SNR)

7. CONCLUSIONS AND FURTHER RESEARCH

Our main goal in this paper was to analyze the usability of a different approach to the classification of signals. This approach uses relations for its decisions, where the relations are learned, rather than defined. The distinguishing characteristic of the formulation of the classification problem described in this paper is the fact that we do not use any fixed set of features, but instead, the systems attempts to find features in the signal that satisfy a given relation. If such a set is found, the decision is positive, i.e., the signal can be classified as being an element of the class associated with the relation. Otherwise, it is a negative decision.

To single out the impact of this approach on the quality of classification we developed two kinds of simulations, one according to our proposed approach, and another one serving as a benchmark. The benchmark simulation also uses relations for decision making, but unlike the approach presented in this paper, it first selects a fixed (such that gives best discrimination power) set of features and then finds a best relation for this set of features. The recognition accuracy based on a fixed set of features does not seem to be a good choice for the scenarios similar to the one presented in this paper due to the fact that the signals in one class were similar in terms of their shape, but were quite different if taken simply as functions of time. For instance, the triangles had their base varying in both size and location, the apex of the triangle varied in both the height and the location with respect to the base. For rectangular signals, the base size and position, as well as the height, were varied.

As can be seen from Figure 4, the classification accuracy for the method proposed in this paper is significantly better than for the benchmark approach. It is uniformly better for all levels of SNR tested in our experiments. We attribute this improvement to the use of relations with variable features, as opposed to using fixed sets of features.

While this tells us that the approach proposed in this paper is worth further study, it also requires more work to make it applicable in practice. First of all, the algorithms investigated in this study exhibit a high level of sensitivity to noise. This could be expected since the algorithms did not use any noise filtering or smoothing, but they rather operated on pure signals. This can be improved using standard filtering and smoothing techniques.

Another direction that should be investigated deeper is the choice of the learning and classification algorithms. In this study, we used exhaustive, non-incremental algorithms. In other words, we searched the whole space of 5-ary relations on the whole learning database. There are various learning algorithms that could give better performance in both time efficiency and classification accuracy.

Additionally, we considered only 5-ary relations, where the magic number "5" was selected arbitrarily. This could be a parameter in the relation learning algorithm and thus it would be selected by the algorithm dynamically, rather than fixed at the outset.

Acknowledgments

This research was partially supported by a grant from the Office of Scientific Researach under contract No: F49620-98-1-0043. We are grateful to Jerzy Weyman and Jerzy Tomasik for helpful conversations and suggestions.

REFERENCES

1. . . C [illegible] *IEEE Transactions on Information Theory*, 38, . . .2:713 718, 1992.
2. B. . D [illegible] *Decision Fusion*. C [illegible], 1994.
3. [illegible] *IEEE Transactions on Aerospace and Electronic Systems*, 29(4):1244 1251, 1993.
4. [illegible] A [illegible] *International Journal of Information Fusion (in print*, 2001.
5. [illegible] *Applied Signal Processing*, 5:152 169, 1998.
6. [illegible] *Machine Learning*, 1:81 106, 1986.

7. [illegible] *Local Feature Extraction and Its Applications Using a Library of Bases.* [illegible] D [illegible], 1994.
8. [illegible] *Distributed Detection and Data Fusion.* [illegible], 1996.