A wavelet packet algorithm for classification and detection of moving vehicles

Amir Averbuch[1] Eyal Hulata[1] Valery Zheludev[1] Inna Kozlov[2]

[1] Department of Computer Science
School of Mathematical Sciences
Tel Aviv University, Tel Aviv 69978, Israel

[2] Electro-Optics Research and Development Ltd (EORD)
Technion, Haifa 32000, Israel

Abstract

In this paper we propose a robust algorithm that solves two related problems: 1) Classification of acoustic signals emitted by different moving vehicles. The recorded signals have to be assigned to pre-existing categories independently from the recording surrounding conditions. 2) Detection of the presence of a vehicle in a certain class via analysis of its acoustic signature against the existing database of recorded and processed acoustic signals. To achieve this detection with practically no false alarms we construct the acoustic signature of a certain vehicle using the distribution of the energies among blocks which consist of wavelet packet coefficients. We allow no false alarms in the detection even under severe conditions; for example when the acoustic recording of target object is a superposition of the acoustics emitted from other vehicles that belong to other classes. The proposed algorithm is robust even under severe noise and a range of rough surrounding conditions. This technology, which has many algorithmic variations, can be used to solve a wide range of classification and detection problems which are based on acoustic processing which are not related to vehicles. These have numerous applications.

bf Keywords: Wavelet packet, acoustic signature, discriminant block, classification, detection, CART.

1 Introduction

Two intrinsically interesting classification/detection problems based on acoustic information are the problems of classification of acoustic signals emitted by different moving vehicles and the detection of the presence of a vehicle in a certain class via analysis of its acoustic signature against the existing database. The problems have an obvious industrial and military application, but, to our knowledge, only a few papers reported some results in this area [13, 10]. The problems are complex because of the great variability in the surrounding conditions of the recorded database: velocity, distances between the vehicles and the receiver, the existence of other vehicles, the roads the vehicles are traveling on, and the background noise, just to name few.

A crucial factor in having a successful classification (no false alarm) is to construct signatures built from characteristic features that enable to discriminate among recorded classes. Multiscale wavelet analysis provides a promising methodology for this purpose. Recently several wavelet based techniques for features extraction were developed. We mention the Local Discriminant Bases algorithm by Saito and Coifman [16, 17] which extends the Best Basis idea [5] to construction of the wavelet packet bases separating classes of signals. The algorithm was applied to classification of some geological phenomena [18]. Another approach to finding most discriminating elements in the classes of signals was presented in [4]. The method named Discriminant Pursuit is related to the Matching Pursuit method by Mallat [15] The techniques in [12] combine two above approaches. However, the above mentioned wavelet based techniques lack translation invariance in the time domain, this is critical for detection **of moving objects**, since misalignment between different signals can generate false results.

Therefore, we developed a new technique which we call *Discriminant Block Pursuit*. The basic assumption is that the acoustic signature for the class of signals emitted by a certain vehicle is obtained as a combination of the inherent energies in a small set of the most discriminant blocks of the wavelet packet coefficients of the signals. It is justified by the fact that each part of the vehicle emits a distinct acoustic signal which in frequency domain contains only a few dominating bands. As the car moves the conditions are changed and the configuration of these bands may vary, but the general disposition remains. Therefore, the blocks of the wavelet packet coefficients, each of which is related to a certain frequency band, are the relevant tool to base the classification on. These blocks contain the distinctive characteristic features. To some extent, our approach is related to [13] where the covariances of the coefficients of the wavelet transform are used for feature extraction. But our techniques are much more flexible and yield rich opportunities for adaptation to the problems are demonstrated to be remarkably efficient even under severe field conditions. Such an approach to features extraction is universal. It is applicable to the analysis of any classes of signals which manifest quasi-periodic structure.

In the final phase of the process, in order to identify the acoustic signatures of the predetermined classes of signals, we used conventional classifiers such as: Linear Discriminant Analysis (LDA) [11] and Classification and Regression Trees (CART) [3]. For classification problems both classifiers produced similar results but for detection, the CART, being properly modified (as explained in section 7), outperformed overwhelmingly the LDA.

The algorithms were tested in a wide series of field experiments. The results demonstrate that certainly the proposed algorithm is robust and the false alarm rate is near zero.

The paper is organized as follows. In Section 2 we provide the necessary information on wavelet and wavelet packet transforms. Then we formulate the approach to solve the problem. In Section 3 we describe the algorithm which is centered around two basic issues: Selection of the discriminant blocks of the wavelet packet transforms and discrimination of signals. Then, we explain its implementation. Section 4 presents the experiments.

2 Formulation of the approach

2.1 The structure of the signals

We assume that the recorded acoustic data generated by the vehicles belongs to one of the classes $C^k,\ k = 1, 2, 3, 4$. These recordings were taken during a wide range of field experiments under

various surrounding conditions. In particular, the velocities of the vehicles and their distances from the recording device were varied. Moreover, the vehicles traveled on either various paved (asphalt) or unpaved roads, or on a mixture of paved and unpaved roads. A series of background recordings were taken, in which the acoustic signals of the vehicles were not present.
Figure 1 shows some portions of the signals emitted by the vehicles from the classes C^k, $k = 1, 2, 3$, with their Fourier transforms.

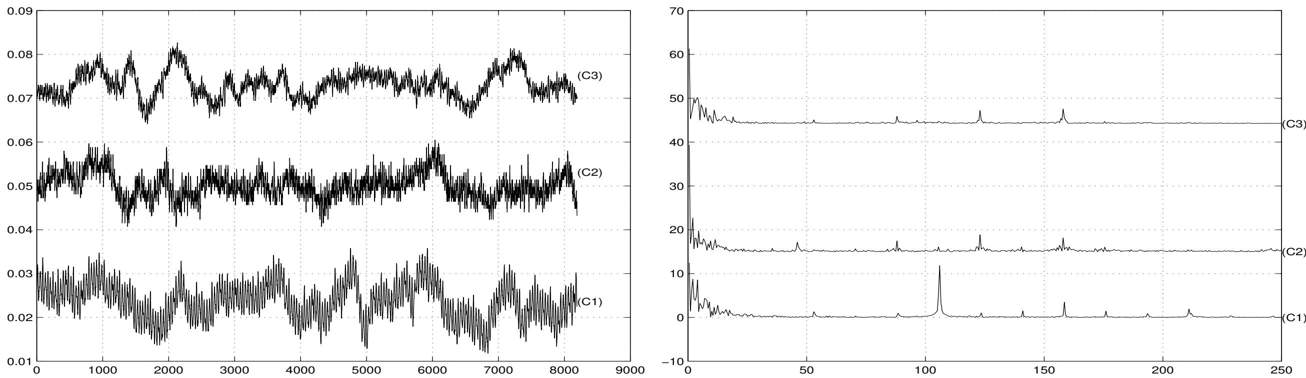

Figure 1: Portions of the signals emitted by vehicles from classes C^k, $k = 1, 2, 3$, (left picture) and of their Fourier transforms (right picture).

Figure 2 displays some portions of a signal from class C^1 and its Fourier transform recorded at different conditions.

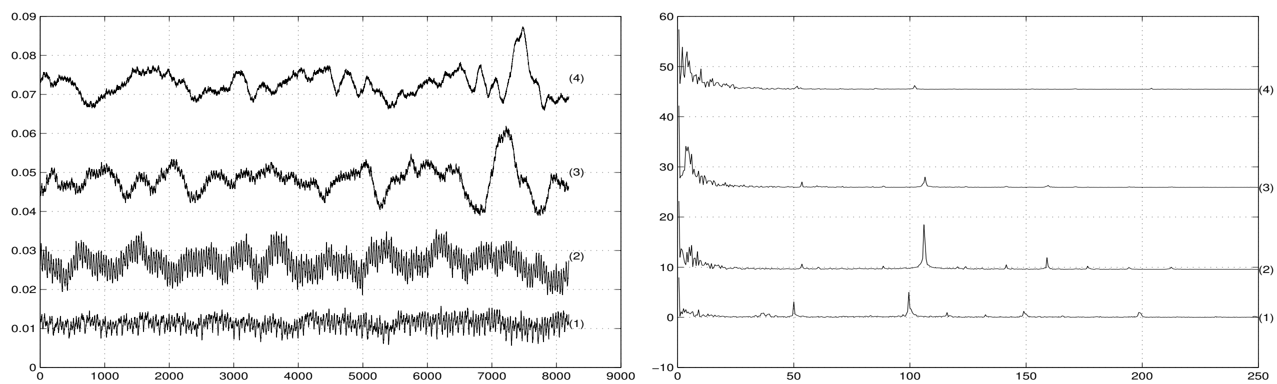

Figure 2: Left picture: sections of class C^1-signal recorded at different conditions: line 1–the vehicle is approaching the receiver; line 2–passing by the receiver; line 3–moving away from the receiver at a moderate speed; line 4–the same but far away from the receiver. Right picture: portions of the Fourier transforms of the signals shown in the left picture.

We realized that even within the same class the signals differ significantly from each other. The same is true for their Fourier transforms. However, there are some common properties to all the acoustic signals that were recorded from moving vehicles. First, these signals are quasi-periodic in the sense that there exist some dominating frequencies in each signal. These frequencies may vary with the change of the motion conditions. However, for the same vehicle these variations are confined in narrow frequency bands. Moreover, the relative locations of the frequency bands are stable (invariant) to some extent for signals that belong to the same class and differ for signals from other classes.

Therefore, we conjectured that the distribution of the energy (or some energy-like parameters) of signals that belong to some class over different areas of the frequency domain may provide a reliable characteristic signature for this class.

2.2 Formulation of the approach

Wavelet packet analysis is a highly relevant tool for adaptive search for valuable frequency bands of a signal or a class of signals. Once implemented, the wavelet packet transform of a signal yields a huge variety of different partitions of the frequency domain. One of the important features of the transform is its computational efficiency. The implementation of an m level transform requires $O(mn)$ operations. Due to the lack of time invariance in the multiscale wavelet packet decomposition, we will deal with the whole blocks of wavelet packet transform rather than with individual coefficients and waveforms. Moreover, following the suggestion in [7] we increase the number of sample signals in the training sets and in the test sets by imposing a comparatively short window on each input signal followed by a shift of this window along the signal so that adjacent sections have some overlap.

General approach:

1. For training we use a set of signals with known membership. From them we select a few blocks which discriminate efficiently between the given classes of signals.
2. We apply the wavelet packet transform on the signal to be classified. We use as its characteristic features which are the normalized l_2 or l_1 norms of the wavelet packet coefficients contained in the selected blocks.
3. Finally, we submit the vectors of the extracted features to one of the classical classifiers. The latter, being appropriately trained beforehand, decides which class this signal belongs to.

3 Description of the algorithm and implementation

3.1 The algorithm

The algorithm is centered around two basic issues:

I. Selection of the discriminant blocks of the wavelet packet coefficients is done according to the following steps:

1. Choice of the analyzing filters.
2. Construction of the training set.
3. Calculation of the energy map.
4. Evaluation of the discriminant power of the decomposition blocks.

5. Selection of the discriminant blocks.

II. Discrimination among the signals.

1. Preparation of the pattern set.
2. Building the CART classification trees.
3. Preparation of the testing set.
4. Making the decision.

Now we present a detailed description for the implementation of the algorithm.

3.2 Implementation

3.2.1 Selection of discriminant blocks

Choice of analyzing waveforms: By now a broad variety of orthogonal and biorthogonal filters that generate wavelet packets coefficients are available. We use 8-th order spline wavelet packets and the Coiflet5 wavelet packets with 10 vanishing moments. These filters reduce the overlapping among frequency bands associated with different decomposition blocks. The empirical analysis suggests to decompose the signals into 5 to 7 levels (scales).

Construction of the training set: Initially, we gather as many recordings as possible for each class C^l, $l = 1, \ldots, L$ which have to be separated. Then we prepare from each selected recording, which belongs to a certain class, a number of overlapping slices of length $n = 2^J$ samples each, shifted with respect to each other by $s \ll n$ samples. All together we produce M^l slices for the class C^l. These groups of slices form the training set for the search of discriminant blocks.

Calculation of the energy map: First, we specified the kind of energy type measure to be used. Typically, we use the normalized l_2 or l_1 norms of the blocks. After the measure has been chosen, the following operations are carried out:

1. The wavelet packet transform is applied up to scale m on each slice of length n from a given class C^l. This procedure produces mn coefficients arranged into $2^{m+1} - 1$ blocks associated with different frequency bands.
2. The slice $A^l(i,:)$ is decomposed. The "energies" of each block are calculated in accordance with the chosen measure. As a result we obtain, to some extent, the distribution of the "energies" of the slice $A^l(i,:)$ over various frequency bands of widths from $N_F/2$ to N_F/m, where N_F is Nyquist frequency. It is presented by an energy vector E^l_i of length $2^{m+1} - 1$.
3. The energy vectors along the training set of the class are averaged:

$$E^l = \frac{1}{M} \sum_{i=1}^{M} E^l_i.$$

   The average energy map E^l of length $2^{m+1} - 1$ indicates how the distribution of the "energies" among various block of the decomposition and frequency bands, respectively, is taking place within the whole class C^l. Similar operations are performed on all the classes C^l, , $l = 1, \ldots, L$.

The left picture in Figure 3 displays a typical energy map for 5 decomposition levels of a two-class problem. Heights of the bars indicate the normalized energy of each of the 63 decomposition blocks.

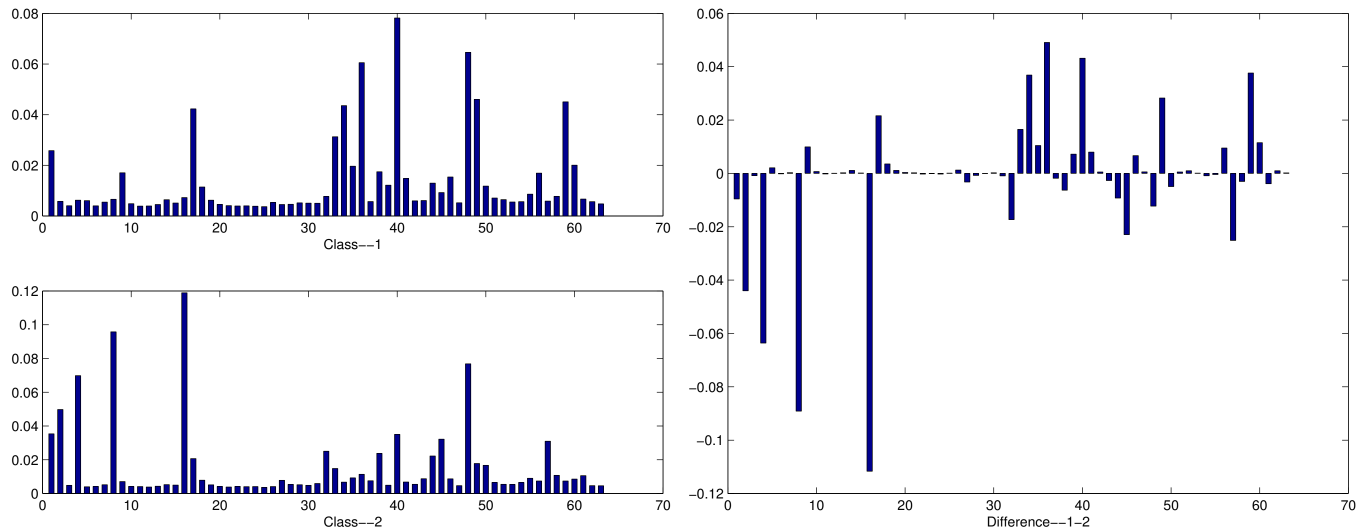

Figure 3: Energy map for 5 decomposition levels of a two-class problem (left picture) and difference of class C^1 and class C^2 maps (right picture). The length of a slice is $n = 1024$ samples.

Evaluation of the discriminant power of decomposition blocks and selection of discriminating blocks The average energy map E^l yields some sort of characterization for the class C^l but it is highly redundant. To gain a more concise and meaningful representation of the class we select the most discriminating blocks. One possible way to do so is the following. First, note that for a two-class problem the difference between two maps provides some insight into the problem (Figure 3 - right). As demonstrated in the figure, the differences for most blocks are near zero. It means that they are of no use for discrimination unlike a few blocks with large values in their differences. Therefore, the term-wise difference (absolute values) of the energy maps serves as the discriminant power map for the decomposition blocks $DP(1,2) = |E^1 - E^2|$. When we have r classes we can find the discriminant power map as the sum: $DP(1,2,...,r) = \sum_{k=1}^{r-1} \sum_{l=k+1}^{r} DP(k,l)$.

Now we are in a position to select a few discriminant blocks which form a sort of signatures for the classes. This was not possible immediately since we are in a situation where the frequency bands of the blocks overlap. For example, the blocks w_2^3 and w_3^3 (see Appendix I) occupy together the same band as the block w_1^2 which is considered as their "parent". If the latter has a strong discriminant power then probably at least one of the "children" blocks has the same. To avoid this frequency overlap we apply the procedure somewhat similar to the *Best Basis Selection Algorithm* [5], [16], [18]. The idea is to compare among the discriminant power of each pair of the "children" with those of their parent. In the case when the discriminant power of the parent exceeds sum of the children powers, the children blocks are discarded and vice versa. As a result we obtain some non-overlapping set of blocks which map the whole frequency domain of our signals, which are referred to as the "most discriminating basis". Typically, this set contains relatively large number of blocks, especially, if the depth m of the decomposition is large. Therefore, we select from the blocks with the highest discriminant factor. Moreover, if we are interested in certain frequency bands, we can select the corresponding blocks.

Another way to select the most discriminating blocks is to implement thresholding i.e. to discard from the discriminant power map the blocks whose power does not exceed some predetermined threshold level. Then, this sparse map is presented to the "best basis" selection unit (to eliminate frequency overlap).

A slightly different approach is possible in a two-class problem if we notice that blocks which gained positive difference values were stronger for the class C^1 and vice versa. Then we could select discriminant blocks separately from the positive and negative sub-maps of the difference map, i.e. separately for each class.

Conclusion: As a result of the operations described above we discover a relatively small set of decomposition blocks such that the distribution of energies amongst them characterize the classes to be distinguished. This part of the investigation is computationally expensive, especially if, for better robustness, large training sets are involved. On the other hand, this task is performed once.

3.3 Classification

Once we have the set of discriminant blocks $B_1, ..., B_t$, we proceed to the classification phase.

Preparation of the reference set. Initially, we chose a number of recordings that belong to the classes C^l, $l = 1, \dots, L$, to be distinguished, from which we form the reference set. These recordings are sliced similarly to that which was used for the preparation of the training set. For a certain class C^l, we form from each selected recording belonging to the class, a number of overlapping slices each of length n. These are shifted with respect to each other by s samples. We assume that there are μ^l slices related to the class C^l that are are gathered into $\mu^l \times n$ matrix a^l, $l = 1 \dots L$.

Then, we apply the wavelet packet transform up to level m on each row $a^l(i,:)$ of this matrix. After the decomposition of the slice $a^l(i,:)$, we calculate only the "energies" of the t blocks $B_1, ..., B_t$ that were selected before. In doing so we obtain the $1 \times t$ vector $V^l(i,:)$ which we regard as a representative of the slice $a^l(i,:)$. The vectors $V^l(i,:)$ form the $\mu^l \times t$ reference matrix V^l of the class C^l. We do the same for all classes C^l, $l = 1, \dots, L$.

The reference sets are used for two purposes: **1.** As pattern sets for LDA. **2.** As training sets for building CART.

The construction of the tree is done by a binary split of the space of input patterns $X \longrightarrow \{X_1 \bigcup X_2 \bigcup \dots \bigcup X_r\}$, so that, once a vector appeared in the subspace X_k, its membership could be predicted with a reasonable reliability. The basic idea with the split is that the data in each descendant subsets is more "pure" than the data in the parent subset. We describe the scheme in more details in Appendix II.

After the construction of the classification tree, with pattern sets for LDA, we are in a position to classify test signals. To do so we must preprocess these signals.

Preparation of the test set. Suppose we are given a signal S whose membership in a certain class has to be established. The algorithm has to be capable to process either a fragment of the recording, or the entire recording, or even a number of recordings. We form from the signal S a number (let it be K) of overlapping slices of length n each, shifted with respect to each other by s samples. All the K slices are gathered into $K \times n$ matrix T.

Each row $T(i,:)$ of this matrix is operated by the wavelet packet transform up to level m. In the decomposed slice $T(i,:)$, we calculate the "energies" of the t blocks $B_1, ..., B_t$ that were selected before. In doing so we obtain the $1 \times t$ vector $W(i,:)$ which we regard as a representative of the slice $T(i,:)$. The vectors $W(i,:)$ form the $K \times t$ test matrix W associated with the signal S.

Making the decision. Once the test matrix W is ready, we present each row $W(i,:)$ of the matrix to two classifiers.

1. LDA calculates the Mahalanobis distances of the vector $W(i,:)$ from L pattern sets associated with the classes C^l, $l = 1, \ldots, L$ and attributes it to the class whose distance was the least.
2. CART uses the tree that was constructed before on the basis of the pattern sets. Once a vector is presented to the tree it is assigned to one of the subsets X_k of the input space X. This determines the most probable membership of the vector.

Then we count the numbers of vectors $W(i,:)$ attributed to each class C^l and make the decision in favor of the class C^γ which gets the majority of the vectors $W(i,:)$. The robustness of the decision is checked by the percentage of the vectors $W(i,:)$ attributed to the class C^γ.

4 Results

We conducted two series of experiments:

1. Classification of signals emitted by three types of vehicles - C^k, $k = 1, 2, 3$.
2. Discrimination between signals emitted by vehicles belonging to the class C^4 and the background.

Our main efforts were channeled to the second problem. We tested various families of wavelet packets and various norms for the feature extraction and various combinations of features presented to the LDA and CART classifiers. The best results were achieved with wavelet packets based on Coiflet5 filters with 10 vanishing moments and splines of order 8. CART classifiers in most experiments outperformed LDA.

4.1 Classification experiments

We had three groups of recordings of acoustic signals emitted by the vehicles $C^k,\ k=1,2,3$. The sampling rate was 4000 samples per second. Each group comprised a number of recordings of signals with length of 65536 samples (about 16-second). One signal in each group was emitted by stationary vehicles and the rest by moving vehicles. The motion dynamics for all three vehicles were similar to each other: they stayed for a while motionless with working engines then started to move, passed by the receiver and went away up to a distance of $\approx 1400meter$. Each group C^1 and C^3 contains 24 16-seconds recordings, but there was a gap in the recording of the C^2 vehicle at distances ≈ 600 to 1000 meters. We displayed in Fig. 1 some portions of the signals emitted by vehicles $C^k,\ k=1,2,3$ and their Fourier transforms.

We processed the signals using the scheme explained above. For the selection of the discriminant blocks we used 12 signals from each group recorded at distances up to ≈ 600 meters. The signal is processed under sliding overlapped windows of size $n=1024$. The window was shifted along the signal with a step of $s=128$ samples. Each window was processed by the wavelet packet transform up to 5th or 6th levels (scales). The best results were achieved when we used spline 8 wavelet packets and either l_1 or l_2 norms as the "energy" measure for the blocks. As a result of the procedures that were described in Section 3.2.1 we selected various sets of discriminant blocks.

The pattern set involved 4 signals from each group chosen from 12 signals which we used before for the selection of the discriminant blocks (distances up to $\approx 600m$). The vectors of l_1 or l_2 norms of the selected wavelet packet blocks of the windows formed three reference matrices of size $2016 \times q$, where q is the number of the blocks in the given set. These matrices were used as an input to LDA and the construction of CART trees.

In the testing phase we employed all the available recordings, i.e. 24 recordings of each class C^1 and C^3 and 18 recordings of the class C^2. We tested independently 192 non-overlapping signals with duration of 2 seconds each that were retrieved from the recordings of classes C^1 and C^3 and 144 recordings of class C^2. In other words, a signal was classified after 2-seconds of listening.

In Fig. 4 we present the results of the two most successful LDA experiments.

The top left picture illustrates the classification rate for the signals of class C^1. Each $*$ in the picture corresponds to a single 2-second signal of class C^1. Let us take a signal S^l from the class C^1. The corresponding $*\ s^l$ is located in position l which is related to the distance of the vehicle from the receiver. Its height, h^l, may range from 0 to 1 and is equal to

$$h^l = \frac{K_c^l}{K^l}$$

where K^l is the total number of vectors $W(i,:)$ associated with the signal S^l (in our case $K^l=120$ for all l), and K_c^l is the number of the vectors $W(i,:)$ attributed correctly to the class C^1. So, if $h^l=1$ then the signal S^l is completely classified. If $0.5 < h^l < 1$ then the number of correct answers prevails the number of wrong ones and the signal is classified. The closer h^l gets to 1 the more reliable the answer. If $h^l=0.5$ then the number of correct answers is equal to the number of wrong ones and the signal is non-classified. And, at last, if $h^l < 0.5$ then the signal is misclassified.

In the right top picture the classification rate for the signals of class C^1 with another block set is presented. The central pair of diagrams presents the classification rate for signals of class C^2, and the bottom pair - for class C^3.

In the experiment depicted in the three left-hand side pictures we used the wavelet packets with splines of order 8. We used l_1 norm as the energy measure. Five discriminant blocks from the 5-th decomposition scale were involved: $w_0^5,\ w_{13}^5,\ w_{12}^5, w_1^5,\ w_7^5$.

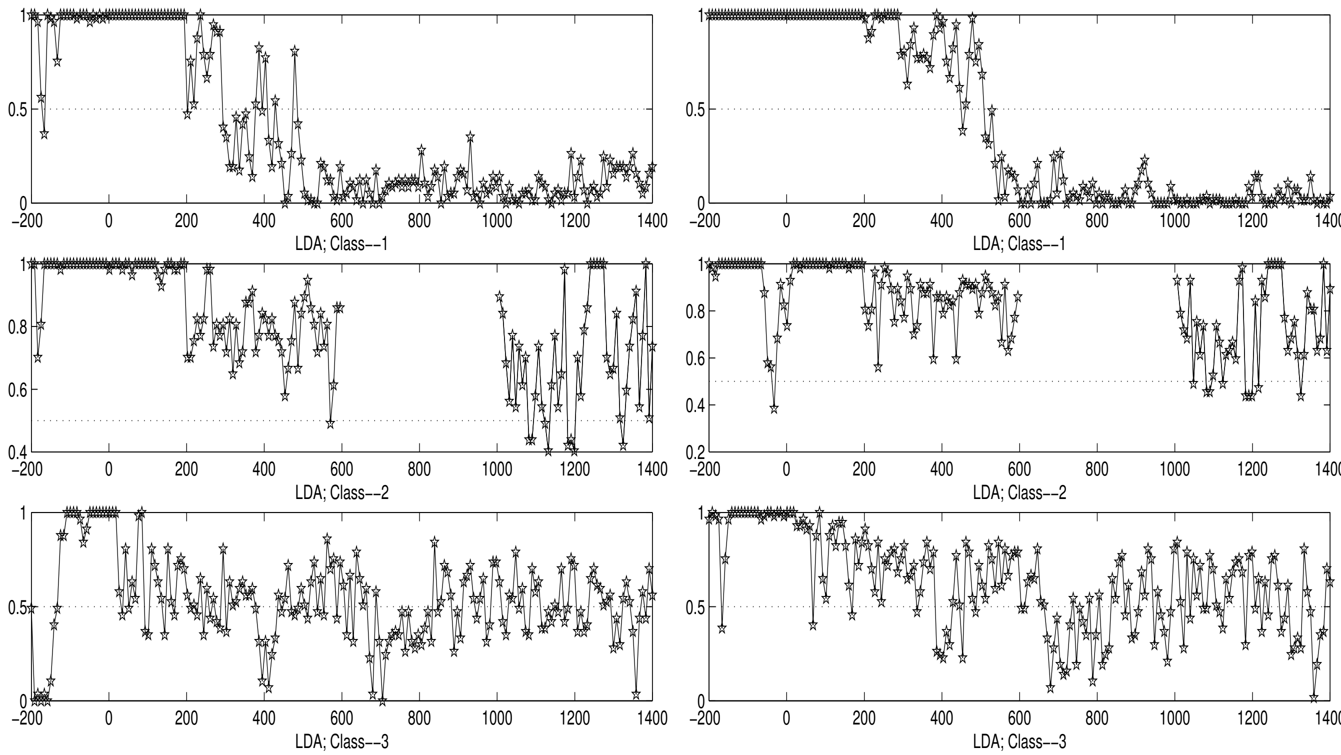

Figure 4: Results from LDA experiments: Left pictures: wavelet–spline8 and energy–l_1 norm were used on 5 discriminant blocks on the 5-th decomposition level. Right pictures: wavelet–spline8 and energy–l_2 norm were used on 9 discriminant blocks on the 4-th and 6-th decomposition levels.

As it is seen from Fig. 4, within the range of 200 meters almost all signals from classes C^1 and C^2 are classified correctly. The majority of C^3 signals are classified as well, but with lesser reliability. Outside this range the classification of C^2 signals work properly up to the ultimate distance of ≈ 1400 meters. Beyond that the classification of C^2 signals fails. It is interesting to note that substantial fraction of the C^3 signals are classified well even when the vehicle is far away from the sensor. C^3 signals are weaker than C^1 and C^2 signals. Moreover, at distances exceeding 800 meters all the signals are hardly audible.

The results of the next LDA experiment are displayed on the right of the triple pictures in Fig. 4. Here more discriminant blocks than before are involved and they were chosen mainly from the 6-th decomposition level, although 2 blocks from the 4-th level were used as well. It means that we used more refined partition of the frequency domain. Moreover, unlike the previous experiment, the energy measure was chosen as the l_2 norm. The blocks involved were: $w_0^6,\ w_6^4,\ w_7^4,\ w_2^6,\ w_3^6,\ w_1^6,\ w_4^6,\ w_5^6,\ w_6^6$. This experiment was more successful. All three groups of signals were classified better than before. In particular, the majority of C^1 signals were classified up to 500 meters distance. The classification of C^3 signals were much more reliable than before. But we must point out that further increase in the

number of the discriminant blocks corrupts rather than improves the quality of the LDA experiments.
In Fig. 5 we display the results of two experiments with the CART classifier. Here the reference matrices were employed for constructing the Classification and Regression Trees. Normally we used more discriminant blocks than with LDA experiments.
In the left from the triple pictures we display an experiment with 10 blocks from the 4-th and 5-th decomposition levels with spline 8 wavelet packets. The l_1 norm energy measure was chosen. The blocks were: w_0^5, w_{13}^5, w_{12}^5, w_7^4, w_1^5, w_7^5, w_2^5, w_6^5, w_2^4, w_8^5.

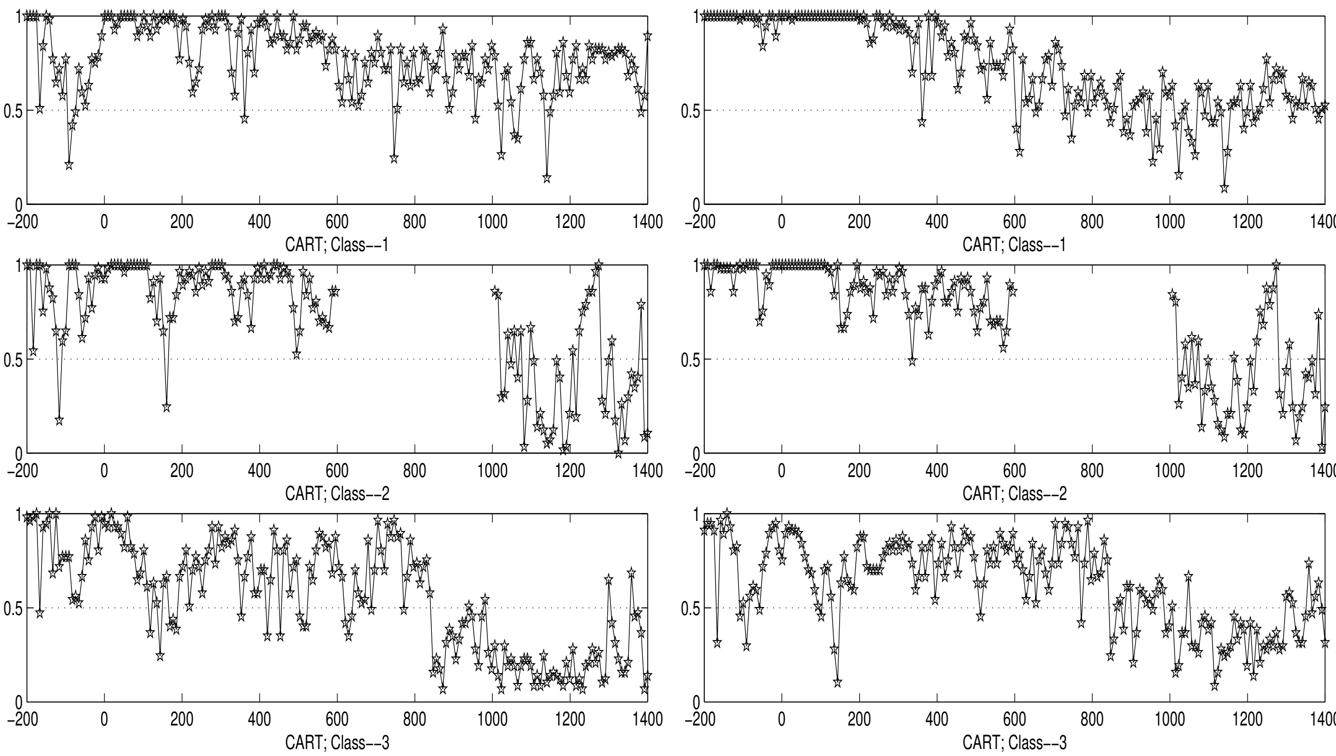

Figure 5: Results from CART experiments: Left pictures: wavelet–spline8 and energy–l_1 norm were used on 10 discriminant blocks from the 4-th and 5-th decomposition levels. Right pictures: wavelet–spline8 and energy–l_2 norm were used on 19 discriminant blocks from the 4-th and 6-th decomposition levels.

The results here are generally better than those from LDA experiments. In particular, it is true for C^1 signals which are classified almost completely up to 1400 meters distance. This success came at the expense of recognizing signals from C^2 and C^3 classes. The quality of the recognized signals from the C^2 and C^3 classes deteriorates at remote distances, although we can still recognize in moderate distances. At 600-800 meters C^3 signals which were not classified by LDA, are classified well now.
In the right picture of the triple we display an experiment with 19 blocks from 4-th and 6-th decomposition levels using spline 8 wavelet packets. The energy measure is the l_2 norm. The blocks were: w_0^6, w_6^4, w_7^4, w_2^6, w_{22}^6, w_3^6, w_{21}^6, w_1^6, w_4^6, w_{23}^6, w_5^6, w_{10}^6, w_6^6, w_9^6, w_8^6, w_7^6, w_{11}^6, w_3^4, w_4^4.
One can observe that up to $\approx$ 600 meters distance the classification became more stable. Beyond

this distance the classification of C^1 signals degrades. In the interval of $600 - 800$ meters the classification of C^3 signals is more successful than in the previous experiment. A few C^3 signals from remote distance are recognized. As for C^2 signals from remote distance, their classification rate was similar to the previous experiment.

Comments

- The classification experiments demonstrated the relevance of our approach to feature extraction.
- The LDA classifier, unlike CART, performed better with a relatively small number of features.
- The use of blocks from the 6-th decomposition level (i.e. from a more refined partition of the frequency domain) improved the performance in comparison to blocks from up to 5 levels.
- We achieved better results with the CART than with LDA. Generally, the CART scheme is more flexible than LDA. It is possible to construct different classification trees with the same reference set. This property provides additional means for the adaptation of the problem to be solved. We will use this opportunity extensively in the detection experiments.

4.2 Discrimination of C^4-signals from background acoustics

If we want to have a robust algorithm that is useful for industrial applications then one of our main goal is to discriminate well between the signals and the background noise. The classification experiments that were described in section 4.1 were an initial step for tackling the problem of detecting a signal emitted by a vehicle of a certain class under whatever surrounding acoustic background the signal of the vehicle was recorded. Formally, this problem is again reduced to classification into two classes:

Class 1 which contains signals emitted by the C^4-vehicles at various velocities, distances and types of roads.

Class 2 which contains signals emitted by other types of different vehicles and pure background noises.

Such a diversity in the structure of Class 2 signals could result in a high rate of false alarms when the standard classification algorithms is being used. To reduce this rate during the selection process of discriminant blocks we choose blocks whose "energy" characterize Class1 signals. We preferred the CART classifier over the LDA classifier . Moreover, we modified the CART algorithm so that in order for a signal to be part of Class 1 it has to pass strict conditions that we imposed upon the classifier. These modifications, which are described in Appendix II Section 7, proved to be very efficient. The false alarm rate was reduced to almost zero whereas by implementation of the standard scheme, it reached up to 64%. We observed that hard-threshold denoising of the signals (see [8]) to be classified improves the results. In all the experiments the energy measure was chosen to be the l_1 norm.
The proposed algorithms were performed on many recorded signals. The algorithms were tested extensively in many experiments with unknown field conditions. These new signals were unavailable during the construction and preprocessing of the algorithms.

As before, we made the decision after a 2-second recording, as was described in Section 3.3. A 2-second signal is labeled as detectable if no less than 80% of its slices $W(i,:)$ belong to Class1. The results of a few experiments are presented in the diagrams. The presentation is similar to that in the previous section where we display in the figure no more than 50 2-second recordings. Moreover, under each diagram we depict the *loudness* level of each recording $S^n = \{s_k^n\}_{k=1}^{8000}$ expressed through root mean square

$$RMS = \sqrt{\sum_{k=1}^{8000} (s_k^n)^2}.$$

Obviously, the *loudness* level corresponds to the distance of the vehicle from the sensor and it strongly depends on the atmospheric and background conditions. It is worth noting that, unlike the classification experiments, the best results were achieved when we used the Coiflet wavelet packets with 5 vanishing moments and the signals were decomposed up to the 7th scale. In all the experiments presented below we used the following common set of discriminant blocks: w_2^6, w_6^7, w_5^6, w_4^6, w_8^6, w_2^7, w_3^5. The same classification tree (CART) was used in all the experiments.

In Fig. 6 we present the detection results of a single vehicle from Class 1 which moves away from the receiver up to a distance of 1600 meters, at 35 km/h on an asphalt road.

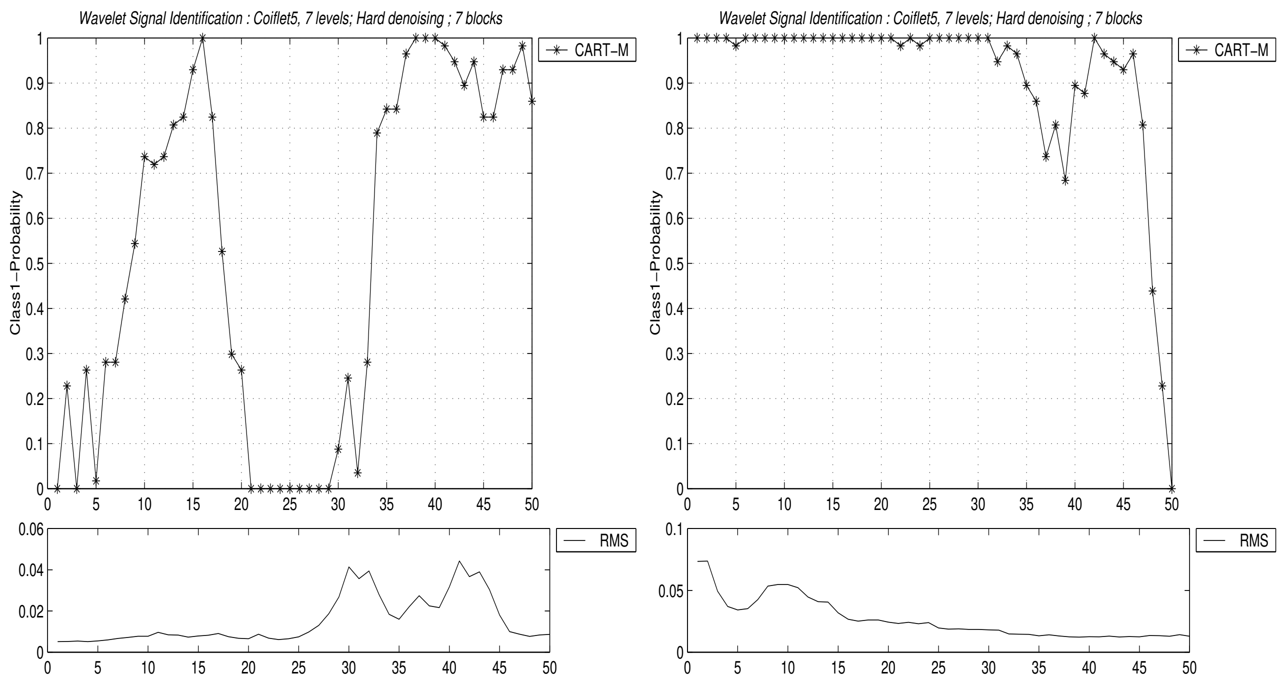

Figure 6: A vehicle from Class 1 is moving away from the receiver up to a distance of 800 meters (left top picture) and from 800 meters to 1600 meters (right top picture) at 35 km/h. Bottom pictures depict the RMS of each 2-second signal. They all traveled on asphalt road. The following parameters were used: Coiflet5 wavelets, 7 blocks, modified CART classifier and every 2 seconds a decision was made.

We can see that a satisfactory detection up to 1400 meters distance was achieved. It is interesting to note that at remote distances (> 300) meters the detection is even more stable than near the receiver.

The drop in the quality of the results in segments 23-28 in the left picture could be explained by a sudden change in atmospheric conditions. We note also that in this experiment the acoustic recording of the vehicle is not relevant after 1100 meters.
In Fig. 7 we present an experiment similar to the previous one, with one difference: the vehicle approached the receiver from a distance of 1800 meters at 40 km/h. The initial distance was $\approx$ 1600 meters. The detection is remarkably stable.

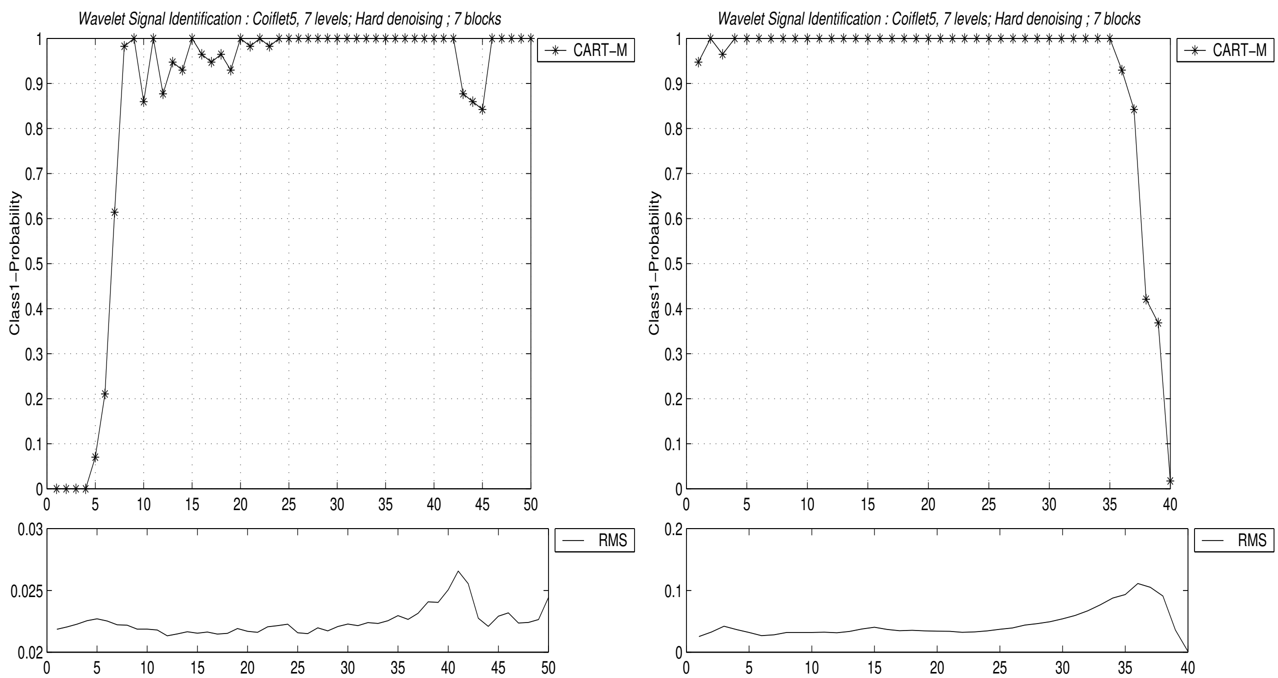

Figure 7: A vehicle from Class 1 is approaching the receiver from a a distance of 1800 meters to 900 meters (left top picture) and from 900 meters (right top picture) at 40 km/h. Bottom pictures depict the RMS of each 2-second signal. They all traveled on asphalt road. The following parameters were used: Coiflet5 wavelets, 7 blocks, modified CART classifier and every 2 seconds a decision was made.

In Fig. 8 the results were taken while the vehicle traveled on the ground instead of traveling on an asphalt road. In the left pictures we present the results when the vehicle is moving away up to 1000 meters from the receiver. In the right picture the results were taken while the vehicle approached the receiver from 900 meters distance. The velocity again was 40 km/h. The range for having a valid acoustic recording was chosen to be 700 meters. One can see that our tool which detected successfully the target on the asphalt, did a good job on the ground road as well.

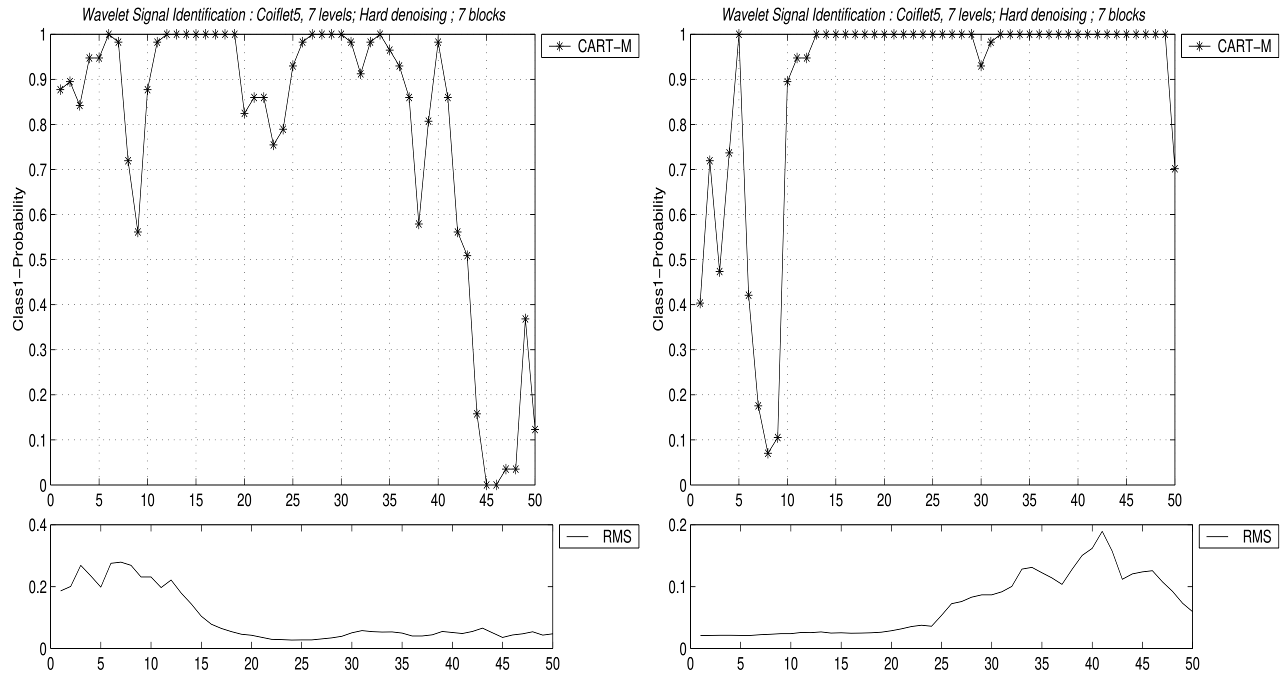

Figure 8: A vehicle from Class 1 is moving away from the receiver up a a distance of 1000 meters (left top picture). The vehicle approached the receiver from 900 meters distance at 40 km/h. Bottom pictures depict the RMS of each 2-second signal. They all traveled on ground road. The following parameters were used: Coiflet5 wavelets, 7 blocks, modified CART classifier and every 2 seconds a decision was made.

In the left pictures of Fig. 9 a more complicated situation is demonstrated. Here the vehicle from Class 1 passed the receiver being accompanied by two vehicles from Class 2. The vehicles followed each another on an asphalt road at 40 km/h. One can see that, despite the strong unfavorable background noises, the vehicle from Class 1 was satisfactorily detected.
In the right pictures of Fig. 9 we address the problem of misclassification or false alarms. While we used the standard CART algorithm, we were unsuccessful to address this problem well. It was resolved after a modification to the algorithm as was explained above was made. The results of the experiment displayed in these pictures is similar to the previous one with the difference that the vehicle from Class 1 is missing from the convoy. The plot with the circles depicts the results of classification with 2-second fragments by the conventional CART algorithm. One can observe that 8 fragments erroneously detected a Class 1 vehicle with probability exceeding 80%. This means that a definite false alarm occurred. When the modified CART was used and this is depicted by the plot with asterisks, the classification results which indicate that it does not belong to Class 1 was with probability exceeding 37%. In the other experiments we also had a very low misclassification rate while using the modified CART.

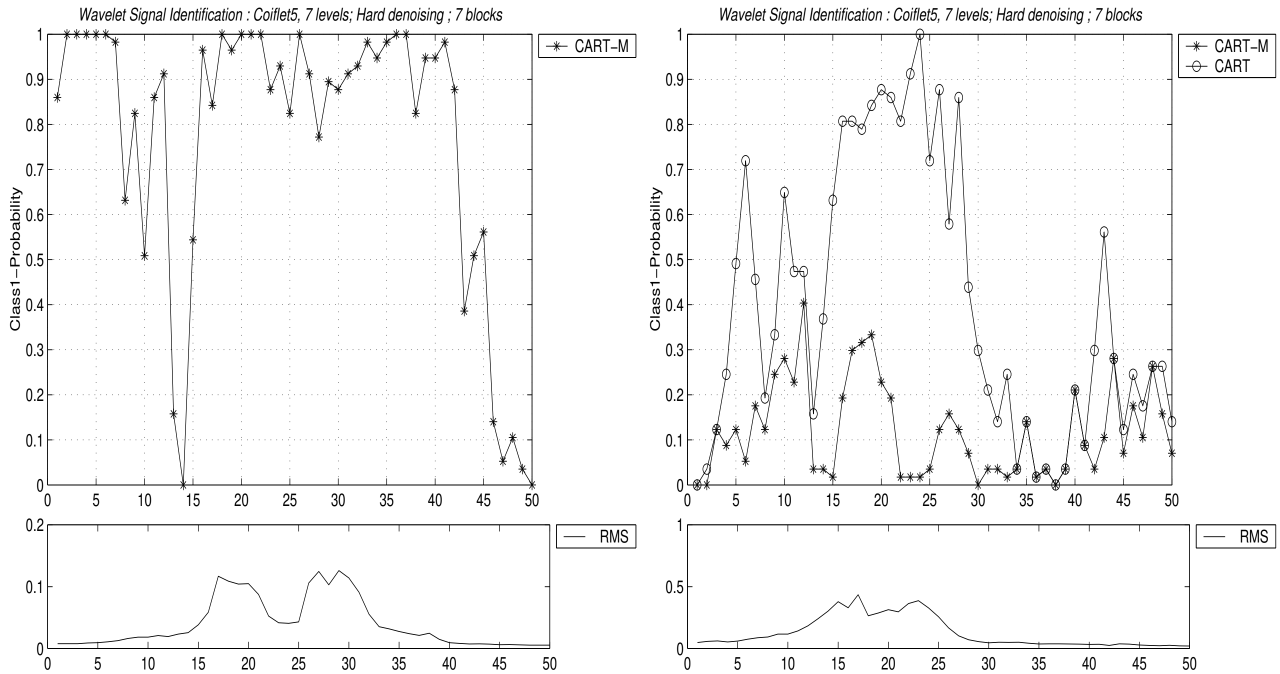

Figure 9: Left pictures: A vehicle from Class 1 is passing the receiver accompanied by two vehicles from Class 2. Right pictures: Two vehicles from Class 2 without the vehicle from Class 1 are passing the receiver. The plot with the circles depicts results of classification of 2-second fragments by the conventional CART algorithm. The plot with asterisks does the same for the modified CART. All is done while the vehicles travel at 40 km/h on asphalt road. Bottom pictures depict the RMS of each 2-second signal. The following parameters were used: Coiflet5 wavelets, 7 blocks. Decision was made every 2 seconds.

The following is a hard problem that remains unsolved. The goal is to detect vehicles of Class 1 moving within a large convoy of other vehicles. The detection must be fulfilled even in the event when some Class 2 vehicles shadow the sought-for targets passing near the receiver. These situations are illustrated in Fig. 10. In the left picture the convoy consists of 10 vehicles and two of them belong to Class 1. Fragments $20 - 40$ are related to the moment when two Class 2 vehicles passed near the receiver shadowing the vehicles of Class 1 to be detected. We can see that for these fragment the detection rate is plunging to near zero. The velocity here was 28 km/h. In the right picture the convoy consists of 6 vehicles and two of them belong to Class 1. It approaches the receiver at a rate of 56 km/h. Fragments $10 - 30$ are related to the moment when two Class 2 vehicles shadowed the vehicles of Class 1 to be detected. It is worth noting that without the presence of this shadowing we sometimes get a satisfactory detection.

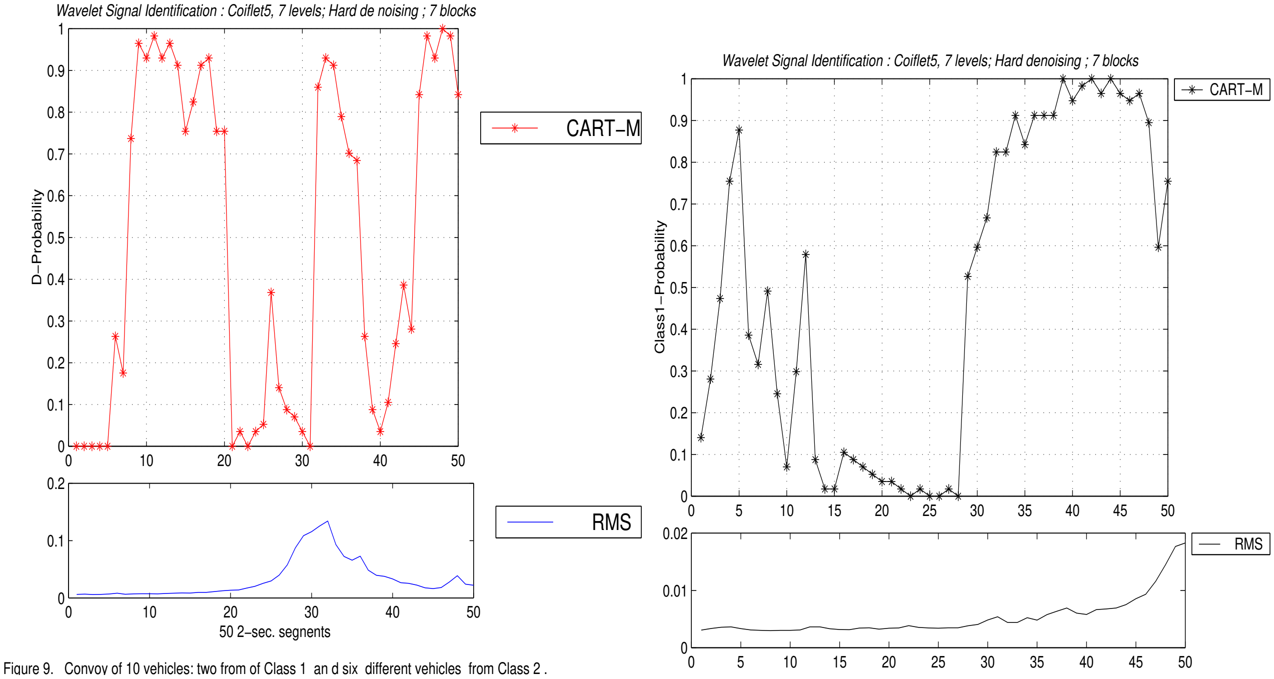

Figure 10: Left pictures: The convoy contains 10 vehicles including two vehicles which belong to Class 1. Fragments 20 ÷ 40 are related to the moment when two Class 2 vehicles passed near the receiver shadowing the vehicles of Class 1 to be detected. The velocity is 28 km/h. Right pictures: The convoy contains 6 vehicles and two of them belong to Class 1. It approaches the receiver at a rate of 56 km/h. Fragments 10 ÷ 30 are related to the moment when two Class 2 vehicles shadow the vehicles of Class 1 to be detected.

5 Conclusions

We presented a robust algorithm that solves two related problems: 1) Classification of acoustic signals emitted by different moving vehicles. The recorded signals have to be assigned to a pre-existing group defined independently from the recording surrounding conditions. 2) Detection of the presence of a vehicle in a certain class via analysis of its acoustic signature against the existing database of recorded and processed acoustic signals.

The second problem was most important because it demonstrates its applicability to the solution to real industrial problems. To achieve this detection with minimal false alarms we constructed the acoustic signature of a certain vehicle using the distribution of the energies among blocks which consist of wavelet packet coefficients. As decision units for Problem 1 we used Linear Discriminant Analysis (LDA) and Classification and Regression Tree (CART) classifiers. For Problem 2 the CART algorithm proved to be much more flexible than LDA. Therefore, for detection we used only the CART classifiers. We developed an efficient procedure for the selection of the most discriminating blocks of wavelet packet coefficients. Moreover, we enhanced the detection abilities of the classifier by a proper modification. Utilizing this we succeeded in reducing the false alarm rate to almost zero for the majority of typical situations. The targets were reliably detected even at large distances from the

receiver when the signals were completely inaudible to the human ear. We got satisfactory detection results for targets moving within a convoy of other vehicles.
But this investigation highlighted some unsolved problems. We want to achieve detection with a very unfavorable signal-to-noise ratio (SNR). Such situations arise, for example, when the target recedes at a large distance or when it is shadowed by other vehicles passing near the receiver. To tackle this problem, we intend in further investigation to devise efficient denoising procedures and to refine the selection of characteristic features of the sought-after signals. For this purpose we will employ the recently developed library of biorthogonal wavelets with strong adaptation abilities [2].
This technology, which has many algorithmic variations, can be used to solve a wide range of classification and detection problems which are based on acoustic processing and, more generally, for classification and detection of signals which have near-periodic structure.

6 Appendix I: Wavelet and wavelet packet transforms

By now the wavelet transform and its extension the wavelet packet transform are widespread and have been described comprehensively in the literature [9], [19], [14]. Therefore, we restrict ourselves to mention only relevant facts that are necessary to understand the construction of the algorithm.
The result of the application of the wavelet transform to a signal f of length $n = 2^J$ is a set of n correlation coefficients of the signal with scaled and shifted versions of two basic waveforms – the father and mother wavelets. The transform is implemented through iterated application of a conjugate pair of low– (H) and high– (G) pass filters followed by downsampling. In the first decomposition step the filters are applied to the signal f and, after downsampling, the result has two blocks w_0^1 and w_1^1 of the first scale, each of size $n/2$. These blocks consist of the correlation coefficients of the signal with 2-sample shifts of the low frequency father wavelet and high frequency mother wavelet, respectively. The block w_0^1 contains the coefficients necessary for the reconstruction of the low-frequency component of the signal. Because of the orthogonality of the filters, the energy (l_2 norm) of the block w_0^1 is equal to that of the component W_0^1. Similarly, the high frequency component W_0^1 can be reconstructed from the block w_1^1. In that sense each decomposition block is linked to a certain half of the frequency domain of the signal.
While block w_1^1 is stored on block w_0^1 the same procedure is applied to generate the second level (scale) of blocks w_0^2 and w_1^2 of size $n/4$. These blocks consist of the correlation coefficients with 4-sample shifts of the two times dilated versions of the father and mother wavelets. Their spectra share the low frequency band previously occupied by the original father wavelet. Then w_0^2 is decomposed in the same manner and the procedure is repeated m times. Finally, the signal f is transformed into a set of blocks $f \longrightarrow \{w_0^m, w_1^m, w_1^{m-1}, w_1^{m-2}, \ldots, w_1^2, w_1^1,\}$ up to the m decomposition level. This transform is orthogonal. One block is remained at each level (level) except for the last one. Each block is related to a single library (waveform). Thus the total number of libraries (waveforms) involved in the transform is $m+1$. Their spectra cover the whole frequency domain and split it in a logarithmic manner. Each decomposition block is linked to a certain frequency band (not sharp) and, since the transform is orthogonal, the l_2 norm of the coefficients of the block is equal to the l_2 norm of the component of the signal f whose spectrum occupies this band.
Through the application of the wavelet packet transform many more libraries of waveforms, namely, 2^j waveforms at the the j–th decomposition level are involved. The difference between the wavelet packet and wavelet transforms begins in the second step of the decomposition. Now both blocks w_0^1

and w_1^1 are stored at the first level and at the same time both are processed by pair of filters H and G which generate four blocks w_0^2, w_1^2, w_2^2, w_3^2 in the second level. These are the correlation coefficients of the signal with 4-sample shifts of the four libraries of waveforms whose spectra split the frequency domain into four parts. All of these blocks are stored in the second level and transformed into eight blocks in the third level, etc. The involved waveforms are well localized in time and frequency domains. Their spectra form a refined partition of the frequency domain (into 2^j parts in the scale j). Correspondingly, each block of the wavelet packet transform describes a certain frequency band. Flow of the wavelet packet transform is given by Fig. 11. The partition of the frequency domain corresponds approximately to the location of blocks in the diagram.

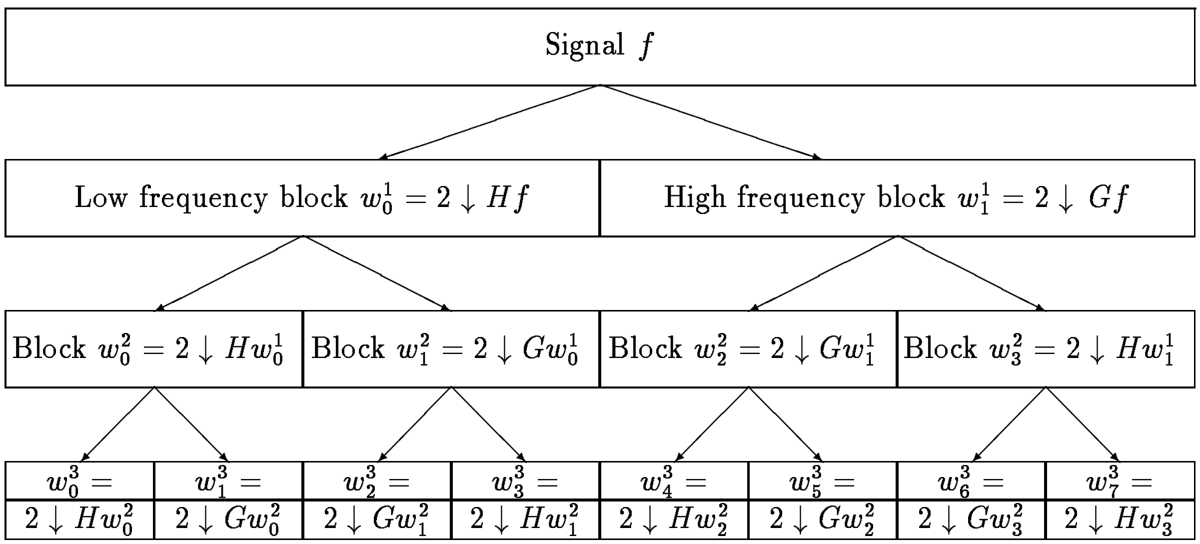

Figure 11: Three levels of wavelet packet decomposition.

There are many wavelet packet libraries. They are differ from each other by their generating filters H and G, the shape of the basic waveforms and the frequency content. In Fig. 12 we describe the wavelet packets after decomposition into three levels generated by the spline of 8-th order. While the splines do not localize well as other wavelets in time domain they produce perfect splitting of the frequency domain.

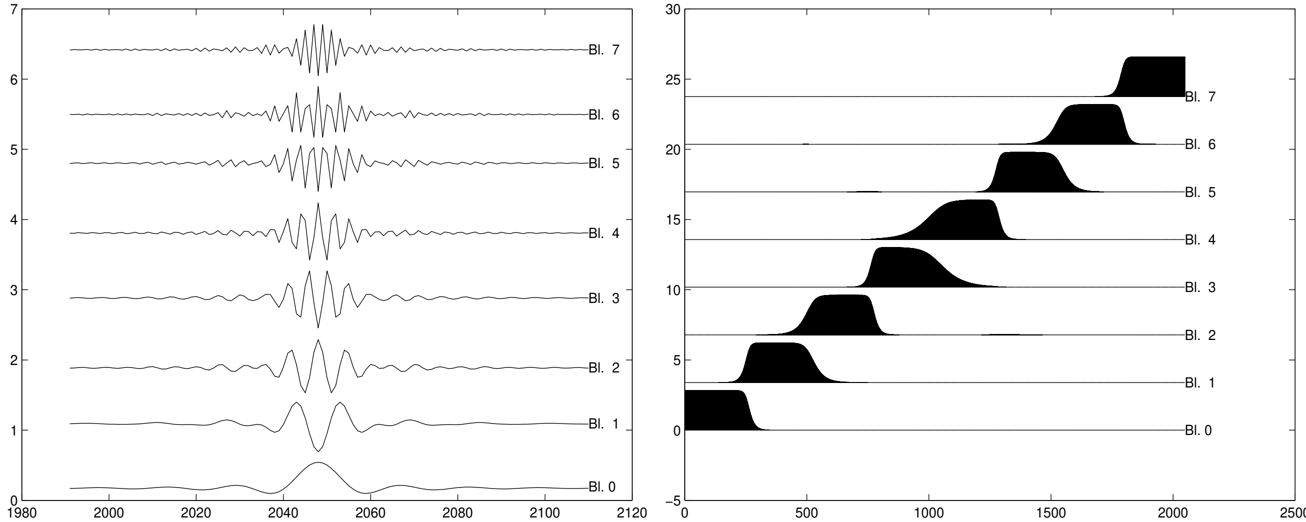

Figure 12: Wavelet packet coefficients after three levels of decomposition generated by spline of the 8-th order and their Fourier spectra. The length of the signal is $n = 1024$.

There is a duality in the nature of the wavelet coefficients of a certain block. On one hand, they indicate the presence of the corresponding waveform in the signal and measure its contribution. On the other hand, they evaluate the contents of the signal inside the related frequency band. We may argue that the wavelet packet transform bridges the gap between time-domain and frequency-domain representations of a signal. As we advance to coarser level (scale) we see a better frequency resolution at the expense of time domain resolution and vice versa. In principle the transform of a signal of length $n = 2^J$ can be implemented up to the J-th decomposition level. At that level there exist n different libraries (waveforms) which are close to sine and cosine waves with multiple frequencies.

7 Appendix II: CART algorithm and its modification

7.1 Outline of the standard scheme

A comprehensive exposition of the CART scheme can be found in [3]. For simplicity of the presentation we consider a two-class classification problem.

Building the tree

The space X of input patterns from the reference set consists of two reference matrices V^l, $l = 1, 2$ of sizes $\mu_l \times n$, respectively. We assume that $\mu_1 = \mu_2$. The i -th row of the matrix V^l is a vector $V^l(i,:)$ of length n representing the signal s_i^l which belongs to the class C^l. In our case, n is equal to the number of discriminant blocks. All row vectors $V^l(i,:)$ should be normalized, i.e.

$$0 \le V^l(i,j) \le 1, \qquad \sum_{j=1}^{t} V^l(i,j) = 1.$$

The tree structured classifier to be constructed has to divide our space X into J disjoint subspaces

$$X = \bigcup_{\nu=1}^{J} X_\nu^t. \tag{1}$$

Each subspace X_ν^t must be "pure" in the sense that the percentage of vectors from one of the matrices V^l, must prevail the percentage of the vectors from the other matrix. (In the original space X both are 50%.)

The construction of the binary tree is started by a split of X into two descendant subspaces:

$$X = X_1 = X_2 \bigcup X_3, \quad X_2 \bigcap X_3 = \emptyset.$$

To do so, CART chooses a split variable y_j and split value z_j in a way to achieve minimal possible "impurity" of the subspaces X_2 and X_3. The split rule for the space X_1 is as follows :

If a vector $y = (y_1, \ldots . y_n)$ satisfies the condition $y_j \leq z_j$, then it transferred to the subspace X_2 , otherwise it is transferred to the subspace X_3. In addition, we divide the subspace X_2 in a similar manner:

$$X_2 = X_4 \bigcup X_5, \quad X4 \bigcap X_5 = \emptyset.$$

The subsequent split variable y_k and split value z_k are selected so that the data in each of the descendant subspaces were "purer than the data in the parent subspace. Then one of the subspaces X_4 or X_5 can be further divided recursively until we reach to the so called terminal subspace X_1^t which is not split further. The decisions whether a subspace is classified as terminal subspace depends on the predetermined minimal "impurity" and the minimal size of the subspace. The terminal subspace X_1^t is assigned to the class C^l, with the probability

$$p_1^l = \frac{m_1^l}{m_1} = \frac{\#\{y \in X_1^t \bigcap V^l\}}{\#\{y \in X_1^t\}},$$

where m_1^l is the number of points in node X_1^t that belongs to class C^l and m_i is the total number of points in the subspace X_1^t. After termination is reached in the subspace X_1^t we return to subspace X_3 which was not split. Similarly, we reach the next terminal subspace X_2^t. We do the same with one of yet non-split subspaces and finally the tree of (1) is constructed. In the terminology of graph theory, the space X is called the root node, the nonterminal and terminal subspaces are the nonterminal and terminal nodes. This process is illustrated in Fig. 13. The terminal nodes are marked as rectangles.

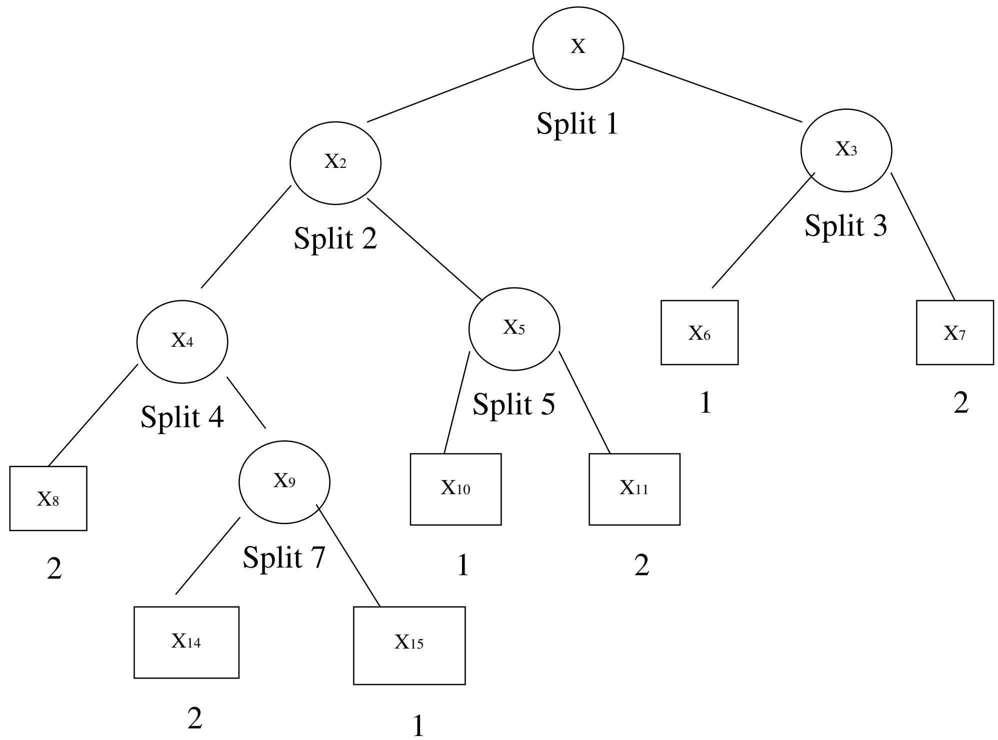

Figure 13: Example of a tree for two class problem. The terminal nodes are indicated by rectangular boxes and are designated by a class label; non-terminal nodes are indicated by circles.

Classification

A vector $x = (x_1, ...x_n)$ is fed to the tree in the first step is assigned to either node X_2 in the case where the coordinate $x_j \leq z_j$ or to node X_3 otherwise. Finally, by checking subsequent split variables, the vector is forwarded into a terminal node X_r^t which is labeled as class C^l, with probability p_r^l.

7.2 CART Modification

The main difference between our specific detection problem and the conventional two-class classification problem lies in the fact that the second (background) class is not properly defined and may contain a huge diversity of signals. While using the standard CART algorithm, we got an unacceptable false alarm rate (FA). For some vehicles from Class C^2, FA ranged up to 64%. To overcome this drawback, we modified the CART algorithm in order to have a detection tool which is based on it. The algorithm has the following steps:

1. Given the space X of input patterns, we build the classification tree as it was explained above (section 7.1). Thus, the space X is split into a set of disjoint terminal subspaces (nodes) $\{X_\nu^t\}$ – (see Eq. 1).

2. The probability value P is predetermined. Usually, the value of P is set to be 0.75. We select from the set $\{X_\nu^t\}$ only the terminal nodes which are assigned to class C^1 with a probability exceeding P. Lets denote such nodes as $\{\Xi^j\}_{j=1}^L$.

3. Let Ξ^j be one of such terminal nodes. Denote $Y^j = V^1 \bigcap \Xi^j$. It means that this is a subset of all vectors $\{y^s = (y_1^s, ...y_n^s)\}_{s=1}^{Sj}$ from the matrix V^1 within the node Ξ^j.

4. We compute
$$\underline{y_i^j} = \min_s y_i^s, \quad \overline{y_i^j} = \max_s y_i^s, \; s = 1, \ldots, S^j.$$
Usually, we do it for all the coordinates $i = 1, \ldots, n$. But sometimes it is reasonable to handle only the most significant coordinates. Similar computation is performed with all the other terminal nodes $\{\Xi^j\}_{j=1}^L$.

5. Let us define the clusters $\{\Gamma^j\}_{j=1}^L$ as follows. We say that a vector $x = (x_1, ...x_n)$ belongs to a cluster Γ^j if all (or significant) coordinates satisfy the conditions
$$\underline{y_i^j} \le x_i^s \le \overline{y_i^j}, \quad i = 1, \ldots, n. \tag{2}$$

6. **Decision rule:** We assign a vector $x = (x_1, ...x_n)$ to Class C^1 if it belongs to one of the clusters Γ^j.

This modified CART algorithm reduced the FA in our experiments almost to zero.

References

[1] A. Averbuch, F. G. Meyer, J.-O. Strömberg, Fast adaptive wavelet packet image compression, to appear in *IEEE Trans. Image Proc.*

[2] A. Averbuch, V. Zheludev *Construction of biorthogonal discrete wavelet transforms using interpolatory splines*, Tel Aviv University, The School of Mathematical Sciences, Technical Report, (1999).

[3] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees, Chapman & Hall, Inc., New York, 1993.

[4] J. Buckheit, D. Donoho, Improved linear discrimination using time-frequency dictionaries,*Proc. SPIE,* **2569**, (1995), 540-551.

[5] R. R. Coifman, and M. V. Wickerhauser, Entropy-based algorithms for best basis selection, *IEEE Trans. Inf . Theory,* **38** (1992), 713-719.

[6] R. R. Coifman, Y. Meyer, M. V. Wickerhauser, Adapted waveform analysis, wavelet-packets, and applications, *In Proceedings of ICIAM'91,* SIAM Press, Philadelphia, 1992, 41-50.

[7] R. R. Coifman, private communication, (1998).

[8] D. Donoho and I. Jonstone, Ideal denoising in an orthonormal basis chosen from a library of bases. *C.R. Acad. Sci. Paris,* Série I, 319, 1317-1322, 1994.

[9] I. Daubechies, Ten lectures on wavelets. SIAM, 1992.

[10] K. B. Eom, *Analysis of acoustic signatures from moving vehicles using time-varying autoregressive models*, Multidimensional Systems and Signal Processing, **10** (1999), 357-378.

[11] R. A. Fisher, The use of multiple measurements in taxonomic problems, *Ann. Eugenics***7**, (1936), 179-188.

[12] Q. Jiang, S.S. Goh, Z. Lin, Local discriminant time-frequency atoms for signal classification, *Signal Processing*, **72**, (1999), 47-52.

[13] R. E. Karlsen, G. R. Gerhart, D. Gorsich and H. C. Choe, Wavelet-based ground vehicle acoustic recognition system, *Proceedings Seventh Annual: "Ground Target Modeling and Validation Conference', August 1996,* 249-256.

[14] S. Mallat, A wavelet tour of signal processing, Acad Press, 1998.

[15] S. Mallat and Z. Zhang, Matching pursuit with time-frequency dictionaries, *IEEE Trans. Sign. Proc.*, **41**(12) (1993), 3397-3415.

[16] N. Saito and R. R. Coifman, Local discriminant bases and their application, *J. Mathematical Imaging and Vision,* **5,** (1995), 337-358.

[17] N. Saito and R. R. Coifman, Improved local discriminant bases using probability density estimation, *Proc. Am. Statist. Assoc., Statistical Computing Section,* **5,** (1996), 312-321.

[18] N. Saito and R. R. Coifman, *Extraction of geological information from acoustic well-logging waveforms using time-frequency wavelets*, Geophysics, **62**, (1997), 1921-1930.

[19] W. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software,* AK Peters, Wellesley, Massachusetts, 1994.