www.czasopisma.pan.pl PAN www.journals.pan.pl
INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2013, VOL. 59, NO. 2, PP. 151–160
Manuscript received March 4, 2013; revised May, 2013. DOI: 10.2478/eletel-2013-0018

A Multistage Procedure of Mobile Vehicle Acoustic Identification for Single-Sensor Embedded Device

Sergei Astapov and Andri Riid

***Abstract*—Mobile vehicle identification has a wide application field for both civilian and military uses. Vehicle identification may be achieved by incorporating single or multiple sensor solutions and through data fusion. This paper considers a single-sensor multistage hierarchical algorithm of acoustic signal analysis and pattern recognition for the identification of mobile vehicles in an open environment. The algorithm applies several standalone techniques to enable complex decision-making during event identification. Computationally inexpensive procedures are specifically chosen in order to provide real-time operation capability. The algorithm is tested on pre-recorded audio signals of civilian vehicles passing the measurement point and shows promising classification accuracy. Implementation on a specific embedded device is also presented and the capability of real-time operation on this device is demonstrated.**

***Keywords*—vehicle identification, acoustic signal analysis, feature extraction, classification, fuzzy logic**

I. Introduction

MOVING object identification is one of many tasks of environment monitoring systems. It finds its uses in civilian and military applications. The civilian applications of moving motor vehicle identification vary from speed limit control to traffic density analysis and traffic behavior prediction. Military uses involve reconnaissance and identification of friendly over enemy craft [1]. The most important aspect of such monitoring systems is real-time computation and timely result processing as the nature of the problem most often implies time-critical operation. Most state of the art systems typically rely on single sensor ultrasonic, acoustic, video, infrared, radar, microwave, magnetic, laser, vibration based, etc. signal analysis, otherwise they employ combinational multisensory detectors [2], [3]. The main advantage of acoustic [4], [5], [6], [7], [8] and video [9] methods lies in the ease of data signal interpretability, i.e., the acquired data is perceptual without additional manipulations.

Video based methods of vehicle identification are generally more effective and robust in changing weather conditions if provided sufficient visibility and illumination. However, the large amounts of video data and significantly more complex pattern search algorithms, if compared to algorithms for one-dimensional data streams, put significant constraints on the possibilities of real-time system implementation. Acoustic systems on the other hand do not rely on visibility factors, yet are sensitive to background acoustic noise variation. Unlike the classical task of distinguishing the incident signal from uniform ambient noise, the task of vehicle identification lies in distinguishing one type of noise (i.e. vehicle-produced sound) from other noises that occur in the environment. Acoustic noise analysis provides the possibility to distinguish well separable classes of motor vehicles, such as passenger cars from large trucks. The acoustic noise patterns of mobile vehicles consist of multiple components [10]. The harmonic nature of the motor noise is, however, seldom present in the civilian vehicle sound pattern due to the fact that motor sounds are well dampened in modern cars. This fact complemented by the Doppler Effect renders the spectral analysis based on fundamental frequency detection (e.g. [11]) ineffective. Instead, parameters of the spectrum overall shape and energy distribution resembling the vehicle noise patterns may be adopted.

This paper considers different methods of digital audio signal analysis, namely the estimation of spectral energy levels and energy envelope, the analysis of several frequency spectrum instantaneous features and spectral pattern matching. The proposed algorithm possesses a hierarchical structure, beginning with the detection of signal perturbation, continuing with the estimation of noise resemblance to those produced by vehicles, and ending with the classification of the detected vehicle. The algorithm is computationally inexpensive and thus is well implementable on embedded devices. We focus on a single-sensor approach in order to reduce the computational load of the algorithm. However, the procedure can be integrated into a more complex system through data fusion for more sophisticated decision-making.

The paper is organized as follows. In Section II, the applied methods of audio signal analysis and audio feature extraction are reviewed in detail. Section III handles the proposed algorithm's procedures and multistage decision-making. In Section IV several computational simplifications are discussed for optimization. In Section V we use two real test signals for experimental verification of the algorithm's identification accuracy and present intermediate and final results of detection and classification. This section of the paper shows that the algorithm is well applicable to the task of identifying motorized vehicles under varying weather conditions. Additionally, in Section VI we present one option of procedure implementation on specific embedded hardware and demonstrate the real-time operation capability on this specific device.

This research was supported by the Innovative Manufacturing Engineering Systems Competence Centre IMECC, co-financed by European Union Regional Development Fund (project EU30006).

S. Astapov is with the Laboratory for Proactive Technologies, Tallinn University of Technology, Ehitajate tee 5, 19086, Tallinn, Estonia (e-mail: sergei.astapov@dcc.ttu.ee).

A. Riid is with the Laboratory for Proactive Technologies, Tallinn University of Technology, Ehitajate tee 5, 19086, Tallinn, Estonia (e-mail: andri@dcc.ttu.ee).

II. Signal Analysis Methods

The audio signal is analyzed in the frequency domain. The frequency domain representation of the signal is achieved by applying a temporal signal decomposing operation, namely the Fourier Transform (FT). Frequency features are less affected by noise than temporal features; also most of the temporal features may be approximated in the frequency domain. Furthermore, the frequency spectrum of a temporal signal frame consists of half as many points as there are in the temporal frame, which is relevant in computation complexity critical systems.

A. The Fast Fourier Transform

The discrete temporal signal is decomposed by the Discrete Fourier Transform (DFT). For a finite duration discrete signal $x(m)$ of length N, the DFT function is

$$X(k) = \sum_{m=0}^{N-1} x(m) \cdot e^{-j\frac{2\pi}{N}mk},\ k = 0, \ldots, N-1. \quad (1)$$

In this manner the transform is performed along two integer dimensions: m and k, i.e. it can be presented as a linear system transformation of complexity $\mathcal{O}(N^2)$. In order to reduce its computation the Fast Fourier Transforms were developed. The proposed system applies a specific implementation of the FFT developed by Frigo and Johnson, called FFTW [12].

The frequency spectrum $[X(0), X(1), \ldots, X(N-1)]$ is symmetrically divided into complex conjugate "positive" and "negative" frequencies, the positive ones residing in the interval $[X(0), \ldots, X(N/2+1)]$ with $X(0)$ being the signal DC component, which is ignored. In order to obtain the absolute amplitude spectrum, the absolute values of this portion of the spectrum are calculated. Thus, abiding the Nyquist – Shannon sampling theorem, the amplitude frequency spectrum of a signal frame of length N consists of $N/2$ frequency components, each of which is multiple to the frequency resolution given by $\Delta f = F_s/N$, where F_s is the sampling rate.

B. Instantaneous Feature Extraction

In order to acquire the specific signal properties, several features are extracted from the amplitude frequency spectrum [13]. These are referred to as instantaneous features due to the fact that they are extracted from every single spectral frame independently, not relying on previous information. The list of features is signal-specific and is formed during the process of test signal analysis in order to distinguish well separable, desirably weakly correlated features, which best indicate the nature of signal fluctuations corresponding to the concerned events. The six spectral features considered in this paper are extracted from the absolute magnitude spectrum frame $|X_t(k)|$ of length , $K = N/2$, $k = 1, \ldots, K$.

Root Mean Square (RMS) Energy of the power spectrum conveys the general spectral energy level:

$$X_{RMS} = \sqrt{\frac{1}{K}\sum_{k=1}^{K} |X_t(k)|^2}. \quad (2)$$

The **band energy** measures the energy of the power spectrum at the i^{th} band and is computed as

$$X_{BE}(i) = \frac{\sum_{l \in S_i} |X_t(l)|^2}{\sum_{k=1}^{K} |X_t(k)|^2}, \quad (3)$$

where S_i is the set of power spectrum samples belonging to the i^{th} band. The bands are chosen according to the Mel-scale denoted by

$$Mel(f) = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right). \quad (4)$$

The Mel-scale is chosen for its increasing spread towards the higher frequencies, which ultimately means that the bands of lower frequencies where most of the spectral energy resides, are shorter than the bands of low-energy higher frequencies. This allows for better distribution of spectral energy by bands.

The **spectral centroid** represents the first central moment of the magnitude spectrum. It is calculated as the frequency averaged over the absolute magnitude spectrum:

$$X_{SC} = \frac{\sum_{k=1}^{K} k \cdot |X_t(k)|}{\sum_{k=1}^{K} |X_t(k)|}. \quad (5)$$

Spectral roll-off measures the frequency below which a certain amount of spectral energy resides. This amount is determined by $TH = [0, 1]$ which is the threshold. For our application we choose it to be equal to $TH = 0.9$ (see Fig. 1).

$$X_{SR} = \arg\max_{p} \left[\sum_{l=1}^{p} |X_t(l)|^2 \leq TH \cdot \sum_{k=1}^{K} |X_t(k)|^2\right] \quad (6)$$

Spectral slope is a measure of spectral energy decrease in the direction of higher frequencies. It is determined by the gradient and y-intersect parameters of a straight line calculated applying linear regression to the magnitude spectrum frame. Hereby, for a set of data points $(k, |X_t(k)|)$, where $k = 1, \ldots, K$, the gradient of the best fitted straight line is denoted as

$$m = \frac{K\sum_{k=1}^{K} k \cdot |X_t(k)| - \sum_{k=1}^{K} k \sum_{k=1}^{K} |X_t(k)|}{K\sum_{k=1}^{K} k^2 - \left(\sum_{k=1}^{K} k\right)^2}, \quad (7)$$

and the y-intersect is denoted as

$$c = \frac{\sum_{k=1}^{K} |X_t(k)| \sum_{k=1}^{K} k^2 - \sum_{k=1}^{K} k \sum_{k=1}^{K} k \cdot |X_t(k)|}{K\sum_{k=1}^{K} k^2 - \left(\sum_{k=1}^{K} k\right)^2}. \quad (8)$$

An example of spectral slope for a magnitude spectrum of length $K = 8192$ is presented in Fig. 1. The overall decline of spectral energy towards higher frequencies defines the parameters of the straight line and not the precise energy distribution in bands.

In the proposed algorithm the RMS energy is used independently. The rest of the considered features are concatenated into a feature vector, which is analyzed during the later stages of classification. A combination of features of different nature

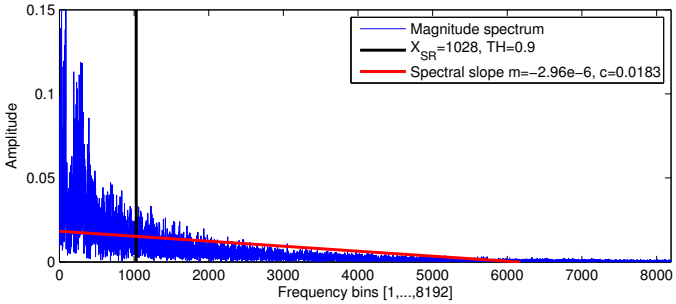

Fig. 1. Spectral roll-off and spectral slope features of an acoustic signal frame.

into a single set may prove harmful in later analysis due to differences of their responsiveness to the incident signal. The method of fuzzy classification, however, overcomes this problem. The issue is further addressed in Section III.

C. Attack Sustain Release Envelope

The process of a vehicle passing the measurement point at a given velocity consists of three stages: approach (spectral energy increases), passing (spectral energy remains stable), retreat (spectral energy decreases). This dynamic pattern is detected by estimating the Attack Sustain Release (ASR) envelope. It is conducted by analyzing the RMS spectral energy (2).

The amount of deviation of RMS energy of the present i^{th} frame $X_{RMS}(i)$ is estimated by the difference between it and the mean value of M previous RMS energy readings. The parameter $\delta \in [0, 1]$ is the lower threshold of energy deviation. RMS energy deviation is coded to three states by the following principle:

$$state_i = \begin{cases} 1, & X_{RMS}(i) > (1+\delta) \cdot mean_{RMS}(i) \\ -1, & X_{RMS}(i) < (1-\delta) \cdot mean_{RMS}(i) \\ 0, & \text{otherwise} \end{cases}, \quad (9)$$

where 1 denotes energy increase, 0 denotes stable energy levels and -1 denotes energy decrease. The mean of M previous energy levels is calculated by

$$mean_{RMS}(i) = \frac{1}{M} \sum_{M}^{j=i-1} X_{RMS}(j). \quad (10)$$

Therefore the transitions $1 \to 0 \to -1$ and $1 \to -1$ are suspected for a car passing event and the quantities of -1, 0, and 1 coded frames denote the lengths of attack, sustain, and release components, respectively.

III. The Hierarchical Algorithm

The proposed hierarchical algorithm, presented in Fig. 2., consists of two independent stages. The hierarchical decision-making scheme (on the left) firstly differentiates relatively loud sounds from mild background noise, secondly it distinguishes vehicle-produced sounds from heavy background noise and lastly estimates the vehicle type from a set of predefined types. This part of the algorithm operates in a frame-by-frame

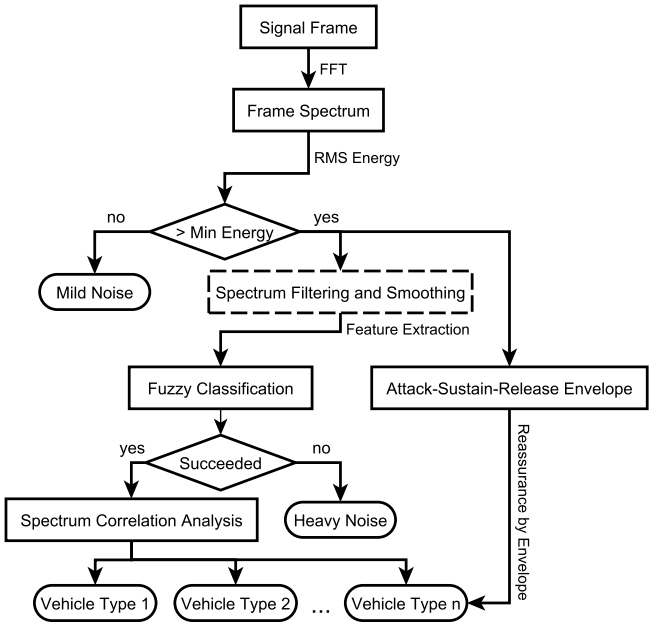

Fig. 2. Block diagram of the proposed hierarchical algorithm for vehicle detection and classification.

manner, computing a single class label per signal frame. The ASR envelope estimation procedure, on the other hand, runs parallel to the decision-making procedure and complements the past frames' classifications with reassurance of positive vehicle-passing event detection. The hierarchy of the algorithm is conditioned by the supremacy of vehicle detection priority over vehicle classification priority, i.e. distinction between vehicle-produced sound and other types of noise is more important than correct vehicle type estimation.

A. Lower Energy Threshold

The first stage of the hierarchical procedure is the estimation of sufficient signal energy. The energy level of a signal frame is calculated and compared to the lower energy threshold, if the threshold is not exceeded, the procedure terminates and the frame is marked as mild noise. The estimation of the lower energy threshold occurs during algorithm parameter estimation by means of test signal analysis. The initial threshold is chosen as the minimal value of RMS energy of all the frames that correspond to vehicle passing instances.

The optional procedure of spectrum filtering and smoothing follows. Digital filtering may improve the Signal to Noise Ratio (SNR) of the spectrum. However, it is effective only in the cases where the spectral band containing the signal is known. In our specific case the vehicle sounds overlap with the background noise and filtering does not improve the classification process. Furthermore, this procedure may corrupt the vehicle acoustic pattern and thus is not applied in our experiments.

B. Fuzzy Classification

The sound pattern of a moving object passing the measuring device is not consistent. Changing signal energy and complex

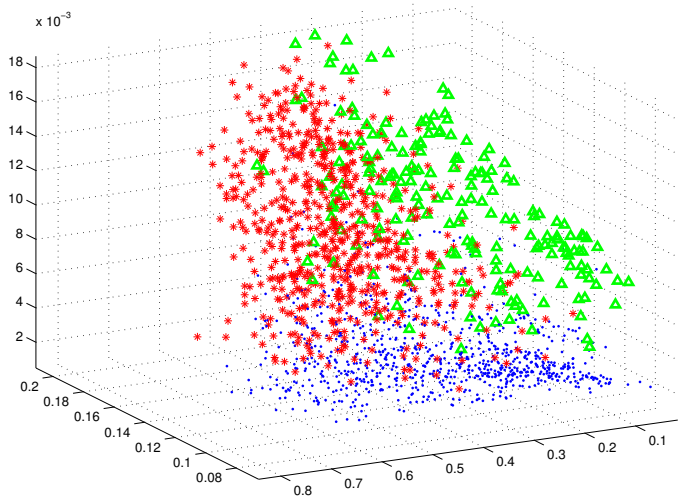

Fig. 3. Three clusters in a three-dimensional feature space. Features represent energy ratios and thus units are not set.

alternations of spectrum shape caused by engine sounds, the Doppler Effect, and also the influence of background noise – all introduce variance to the spectral features of the signal. It is because of this variance why the feature vectors (of length L) corresponding to an event under the same class label form a cluster in an L-dimensional space. Because of the high level of variance among the samples, the clusters can be rather complex shaped (Fig. 3). The separation of samples into a specific class/cluster is carried out by a fuzzy classifier derived by a heuristic training procedure [14]. The classifier operates by applying fuzzy inference to an input feature vector. The algorithm is relatively lightweight if L does not exceed 20-30, which is one of the reasons of applying spectral features instead of the whole spectral vector.

The fuzzy classifier that classifies a feature vector $X = [x_1, \ldots, x_L]$ of length L by assigning to it one of T different discrete valued labels, consists of R rules of the following structure:

$$\begin{array}{c}\text{IF } x_1 \text{ is } A_{1r} \text{ AND } x_2 \text{ is } A_{2r} \text{ AND } \ldots \text{ AND } x_L \text{ is } A_{Lr} \\ \text{THEN } y \text{ belongs to class } c_r\end{array}, \tag{11}$$

where A_{ir} is the linguistic term of the i^{th} input (i.e., feature vector element, $i = 1, \ldots, L$) associated with the r^{th} rule and $c_r \in (1, \ldots, T)$ is the class label assigned by the r^{th} rule.

The class label is assigned in a winner-takes-all manner by specifying the rule with the highest degree of activation

$$y = c_r, \arg\max_{1 \le r \le R} (\tau_r), \tag{12}$$

where τ_r is the activation degree of the r^{th} rule:

$$\tau_r = \bigcap_{i=1}^{L} \mu_{ir}(x_i), \tag{13}$$

where μ_{ir} is the MF corresponding to the linguistic term A_{ir}.

Classifier training consists of estimating the parameters of these MFs. For the implementation of the classifier at hand triangle-shaped MFs are used:

$$\mu_{ir}(x_i) = \begin{cases} \frac{x_i - a_{ir}}{b_{ir} - a_{ir}}, & a_{ir} \le x_i \le b_{ir} \\ \frac{c_{ir} - x_i}{c_{ir} - b_{ir}}, & b_{ir} < x_i \le c_{ir} \\ 0, & (x_i < a_{ir}) \vee (x_i > c_{ir}) \end{cases}, \tag{14}$$

where a_{ir} and c_{ir} locate the base of the triangle and b_{ir} locates the peak. The training is performed using a set of reference feature vectors for which a class label is provided manually. This is done during system off-line tuning. The procedure consists of the following steps:

1) The set of vectors is partitioned into R subsets S_j, $j = 1, \ldots, R$, each consisting of P_j vectors of the same class.
2) The parameters of the MFs are calculated as $a_{ir} = \min_{k \in S_j}(x_i(k))$, $c_{ir} = \max_{k \in S_j}(x_i(k))$, $b_{ir} = \frac{1}{P_j}\sum_{k \in S_j} x_i(k)$, $i = 1, \ldots, L$.
3) The base of each MF is slightly enlarged (by 1% in our case) to give non-zero membership values to the training samples located at the edges of multidimensional space clusters [14].
4) The established MFs are added to the classifier rule-base defined by (13).

If the clusters do not separate naturally in the feature space (which is often the case), the extracted rules are bound to have a high degree of overlap (Fig 4, upper subplot). Note that because of (12), the rules compete for the samples and those samples of a class that are at a sufficient distance from the related rule center, will receive a higher activation degree (13) and consequently, the classification decision from a neighbouring rule by what they lose the connection to the rule they were originally assigned to. In such a case, it makes sense to readjust the MF parameters by excluding the lost samples from corresponding S_j and applying the tuning procedure again. Quite often this ignites a minor chain reaction because the updated rules are inclined to lose additional samples to neighbouring rules and we need to readjust them again. In the end, however, what we obtain is a classifier with much more compact rules and MFs (Fig. 4, lower subplot). Moreover, usually this comes at no loss of classification accuracy.

Note, that the classifier that employs triangular MFs cannot operate on samples that fall beyond the rule borders specified by the MF base parameters. This can be fixed, if desired, by replacing the triangular MFs with near-equivalent Gaussian curves defined as

$$\mu_{ir}(x_i) = \begin{cases} \exp\left\{-\frac{(x_i - b_{ir})^2}{2\cdot(0.4247\cdot(b_{ir} - a_{ir})^2)}\right\}, & x_i < b_{ir} \\ \exp\left\{-\frac{(x_i - b_{ir})^2}{2\cdot(0.4247\cdot(c_{ir} - b_{ir})^2)}\right\}, & x_i \ge b_{ir} \end{cases}. \tag{15}$$

While this improves the ability to properly classify the unseen samples, performance of the classifier first and foremost depends on the quality of the training data set. The reference features must be chosen with moderate amounts of background noise. Very noisy reference features will most likely produce large, sparse and heavily overlapping clusters dependent on the stationary properties of this particular noise. On the other

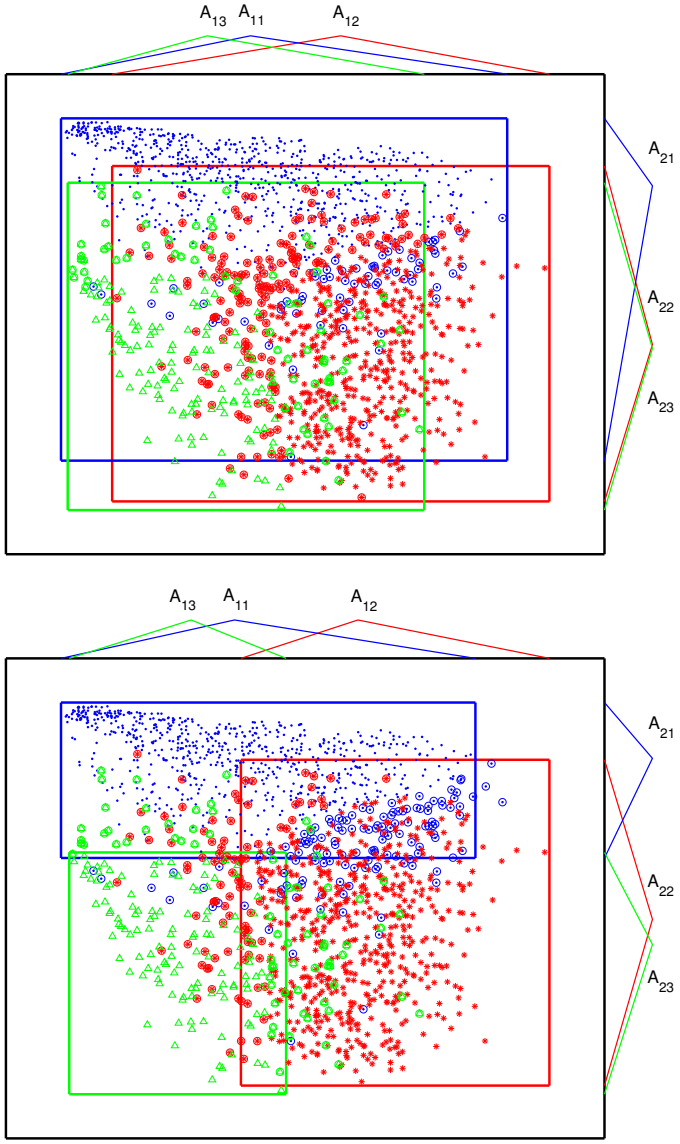

Fig. 4. Initial (upper) and refined (lower) representation of three clusters by triangular MFs.

hand, increasing the size of the reference feature vector will provide more information on the distribution of points in the feature space thus allowing for more efficient MF parameter tuning.

In the proposed hierarchical algorithm, the heuristic fuzzy classification may be applied in two different manners. First, a general cluster corresponding to all vehicle types in question may be estimated and the resulting classifier is used for pure detection purposes, just to distinguish all vehicle produced noise from ambient noise. Final classification is then performed by correlation analysis. Alternatively, a separate cluster for each vehicle class is built and the classifier is applied for specific vehicle type estimation. Here both the fuzzy classifier and the correlation analysis produce separate class estimates, which may reinforce each other. In Section V both methods are used during hierarchical algorithm testing.

C. Correlation Coefficient Analysis

The final stage of vehicle identification is the correlation analysis between the unknown amplitude spectrum vector and the reference spectrum vectors, each corresponding to a single vehicle type class. For a more rigorous classification, several reference vectors per class may also be used. Correlation coefficients are simple and effective metrics for similarity estimation, however, this method is very susceptible to noise. A spectrum of loud background noise may correlate to any of the reference spectra enough to receive incorrect classification. Application of the fuzzy classifier in the previous stage of the algorithm relieves this problem.

During correlation analysis, the correlation coefficients between an unlabeled spectrum vector $x = |X_t(k)|$ and C reference vectors of length K, $r_i = [r_i(1), \ldots, r_i(K)]$, $i = 1, \ldots, N$, are calculated using the following equation:

$$\rho_i = \left[K \sum_{k=1}^{K} x(k) \cdot r_i(k) - \sum_{k=1}^{K} x(k) \sum_{k=1}^{K} r_i(k)\right] / \left[\sqrt{K \sum_{k=1}^{K} x(k)^2 - \left(\sum_{k=1}^{K} x(k)\right)^2} \times \sqrt{K \sum_{k=1}^{K} r_i(k)^2 - \left(\sum_{k=1}^{K} r_i(k)\right)^2}\right] . \tag{16}$$

The correlation is defined on the interval , $-1 \leq \rho_i \leq 1$, -1 meaning total inverse correlation, 0 specifying uncorrelated patterns, and 1 meaning total direct correlation. The class label corresponding to the reference vector of maximum correlation is declared the winner:

$$y = \underset{1 \leq i \leq C}{\arg\max} \left(\rho_i\right) . \tag{17}$$

D. Reassurance by ASR Envelope

As it was mentioned earlier, the detection of the ASR dynamic of signal energy complements the past identification results. If the ASR pattern is detected, a notification is generated and presented along with the final class estimate. The class labels generated during the period of the detected ASR are inspected for the most frequent one (mode in statistical sense), which is presented in the notification. This reduces inconsistencies in the series of class estimates, e.g. when the vehicle type cannot be clearly classified. Additional restrictions may also be applied to the ASR envelope detection. If the potential velocity of the moving object is known, the lower and upper bounds for the attack, sustain or release components may be specified, so the detection is invalid if these restrictions are not met. For example, if the vehicles are known to stop at the measurement point, the expected values of the sustain component have to be large in order to not confuse this stop with multiple vehicles. On the other hand, for quickly passing vehicles on a highway the ASR components are expected to be short.

IV. Algorithm Complexity Minimization

The most time consuming operations of the procedure are feature extraction and correlation coefficient calculation

due to a large number of lengthy vector summations. To reduce the number of summations several feature extraction techniques were specifically chosen with similar summands. Analyzing equations (2), (3), (5) – (8), the repeating elements are $\sum_{k=1}^{K}|X_t(k)|$, $\sum_{k=1}^{K}|X_t(k)|^2$ and $\sum_{k=1}^{K}k\cdot|X_t(k)|$, the first two of which are also present in the correlation calculating equation (16). Computing these sums only once and minimizing the number of cycles during feature extraction greatly reduces the number of overall operations.

Equations (7) and (8) may be further simplified if k is taken as an integer vector index of the corresponding frequency component. The closed form for the sum of K first successive integers is equal to

$$\sum_{k=1}^{K}k=\frac{1}{2}K\left(K+1\right), \tag{18}$$

and the sum of squares of K first successive integers is

$$\sum_{k=1}^{K}k^2=\frac{1}{6}K\left(K+1\right)\left(2K+1\right). \tag{19}$$

Even if k is chosen non-integer, the sums of the resulting recurrences may still be evaluated [15], however these closed forms will definitely require more computations than it is needed for (18) and (19).

Calculating the sums of reference vectors and the sums of squared reference vectors only once during the off-line stage of algorithm parameter specification turns (16) to a more lightweight equation with only one specific summation, which must be performed for each correlation coefficient calculation: $\sum_{k=1}^{K}x(k)\cdot r_i(k)$.

Using a power of two as the signal frame length also reduces computation complexity. FFT operation is optimized for frame lengths multiple to powers of two, also in this case many multiplications and divisions are replaced by simpler and faster bitwise arithmetic shifts.

V. Algorithm Testing Results

The performance of the algorithm is tested on real signals acquired in an open environment. The signals are manually analyzed prior to algorithm accuracy evaluation. This is done in order to estimate the number of classes, that are used in the algorithm and to assign reference class labels to every frame. Each signal is divided into a training and test portion. The partitioning is performed so, that the event of each class occurs at least once in each portion. In our experiments we divide the signals equally, half of the signal is used for training and another half for testing. The training portion is used for fuzzy classifier training and for choosing reference vectors for the correlation analysis procedure. For the fuzzy classifier training, the features are extracted from every frame of the training signal and gathered to the training dataset. For correlation analysis, several spectral vectors corresponding to different classes are manually chosen. The test portion of the signal is used for the estimation of detection and classification accuracy.

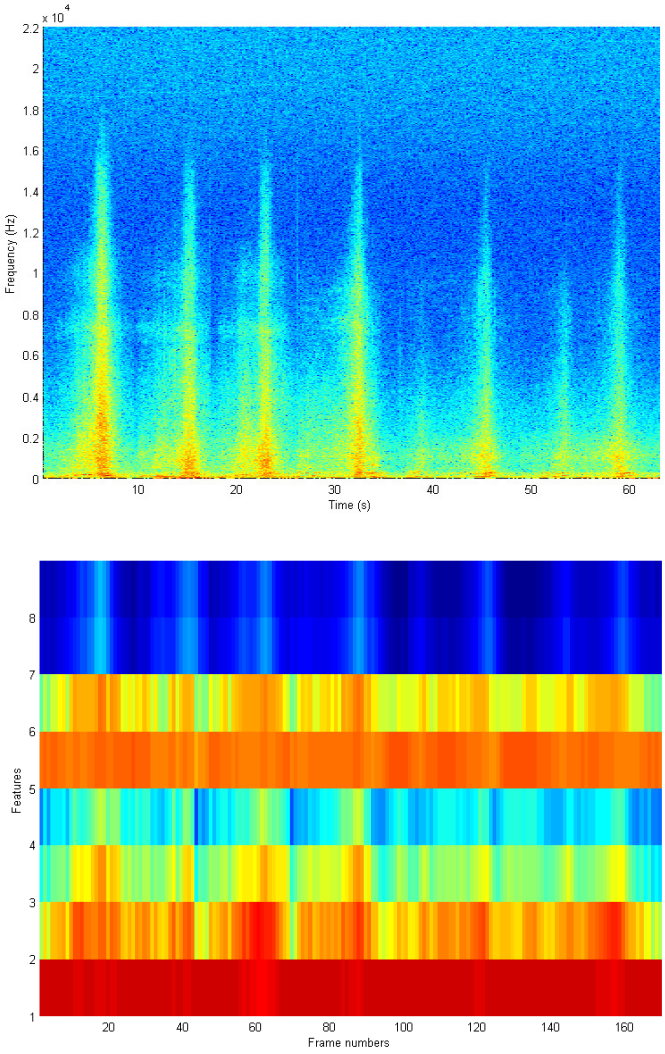

Fig. 5. Spectrogram (upper) and extracted features (lower) of the first experimental signal.

A. First Test Signal

The first test audio signal was measured using a Shure SM58 microphone and a Roland Edirol UA-25EX audio signal processor at 44.1 kHz sampling rate in mono channel mode and saved in 16-bit Waveform Audio File (WAV) format. For the acquisition of the test signal, a microphone was placed at an empty parking lot and two cars (Mercedes S320 and Mazda MX-5) were in turn passing the microphone stand at a speed of 35 – 45 km/h at the passing point, starting to accelerate from a distance of approximately 40 meters. Each car passed the microphone three times: the Mercedes first three times and the Mazda three times afterwards. The sounds were acquired during summer time in mild weather conditions, thus ambient noise levels were relatively low. The signal's spectrogram is presented in Fig. 5. Six instances of passing car sounds are clearly visible.

For testing, the frame length of $2^{14}=16384$ samples is chosen, which corresponds to 0.3715 seconds at a sampling rate of 44.1 kHz. The signal feature vector comprises of eight features: four band energy features (four bands of 1-824, 824-

2616, 2616-6514, 6514-15000 Hz), spectral centroid, spectral roll-off and spectral slope:

$$X = [X_{BE}(1), \ldots, X_{BE}(4), X_{SC}, X_{SR}, m, c]\,. \qquad (20)$$

These features per every signal frame are presented in Fig. 5. Because the SNR is high the features corresponding to car passing events are clearly visually identifiable. For classification a total of 2 class labels are used: 1 for Mercedes and 2 for Mazda. The reference spectral vectors used in correlation analysis are estimated by averaging several spectra of sounds produced by vehicles of the same class, one reference vector per class is applied.

The results of algorithm testing are presented in Fig. 7. The general results are satisfying – every vehicle is detected and successfully classified. As it can be seen in the second and third subplots of Fig. 7, each vehicle passing instance ASR envelope is correctly detected. Though it can be noticed that approximately on the 107^{th} frame the ASR dynamic is falsely detected, the energy of the signal is below the threshold and the fuzzy classifier gives no classification decision, consequently the detection does not occur. Also approximately at the 145^{th} frame the ASR dynamic is present, but is not detected, as for the known vehicle speed corridor the attack and release components of the envelope are set to be no less than 2 frames in duration for the dynamic to be detected. For this signal the fuzzy classifier is trained for detection purposes, i.e. all car types vs. ambient noise. The detection decisions of the fuzzy classifier concur with the ASR envelope.

The heuristic fuzzy algorithm, trained to identify the general vehicle feature space, succeeds in doing so for the majority of signal frames thus allowing the correlation coefficient calculation procedure to analyze only the frames corresponding to vehicle pass time intervals. The fourth subplot of Fig. 7 shows that the correlation coefficient values are unreliable during the periods between vehicle passing instances, during these instances, however, they become more separate, indicating an obvious leader.

B. Second Test Signal

The second test signal was acquired using a condenser microphone Sennheiser KE 4-211-2 and an embedded computing device Gumstix Overo Water. The signal was also sampled at 44.1 kHz mono channel mode and saved to a 16-bit WAV file. Signal acquisition was conducted at a lively two-lane highway during dense traffic in late fall under heavy wind and light rain. Consequently the noise levels in this signal are quite high. The spectrogram, presented in Fig. 6., confirms this. Ambient noise from wind and rain pollutes the whole frequency band, unlike in Fig. 5. Dense traffic results in vehicle passing instances being much less visually separable.

The frame length was chosen the same as for the first test signal. Two vehicle classes were chosen: 1 for passenger cars and 2 for trucks and busses. Feature vectors comprise of eight features, which are the same as for the first signal, except the bands for the band energy features are less spread: 220-818,

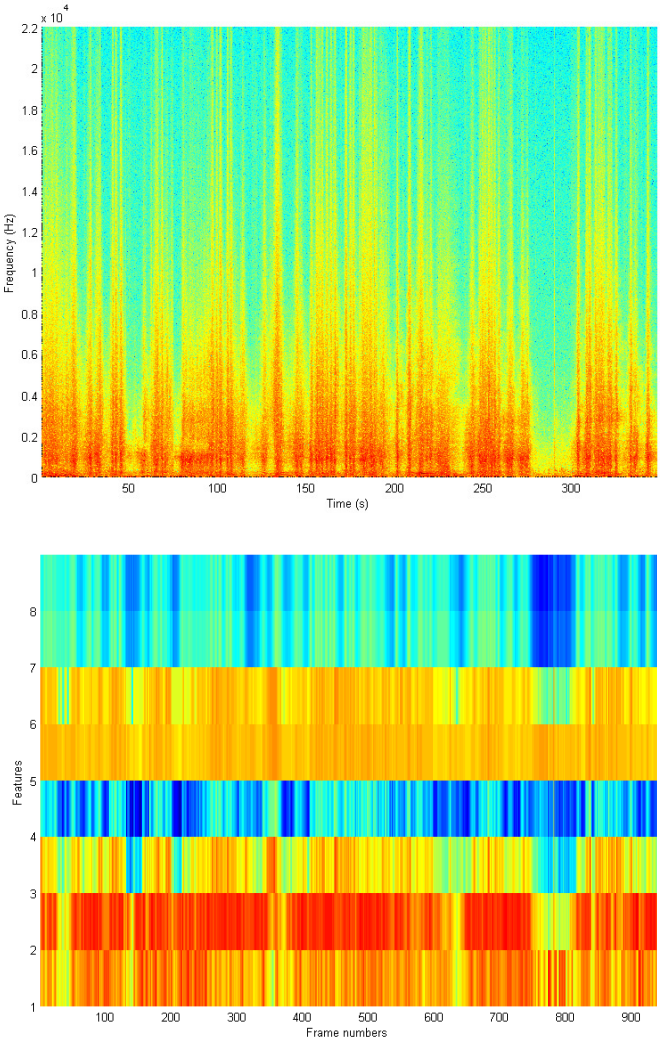

Fig. 6. Spectrogram (upper) and extracted features (lower) of the second experimental signal.

818-2592, 2592-6438, 6438-14780 Hz. Fig. 6. presents these feature vectors. Low SNR makes the features corresponding to vehicle passing instances visually almost unidentifiable. For the derivation of reference spectral vectors, the same technique as for the first signal is used.

The results of signal analysis are presented in Fig. 8. As the time intervals between car passes are very short and often non-existent altogether, reference class labels, which are also used during fuzzy algorithm training, are introduced in the first subplot. The intermediate results are, on the other hand, not presented due to possible problems with readability. The results are as follows: out of 46 instances of class 1 vehicles, 37 were successfully detected and classified, 5 were undetected and 4 were confused with class 2; for 11 instances of class 2 vehicles, 9 were correctly classified, 1 was not detected and 1 confused with class 1. Thus the classification accuracy for class 1 vehicles is 80.43% and for class 2 – 81.82%. Considering the harsh environmental conditions, the overall detection and classification accuracy is admissible.

The main problems causing the lowered classification accuracy are:

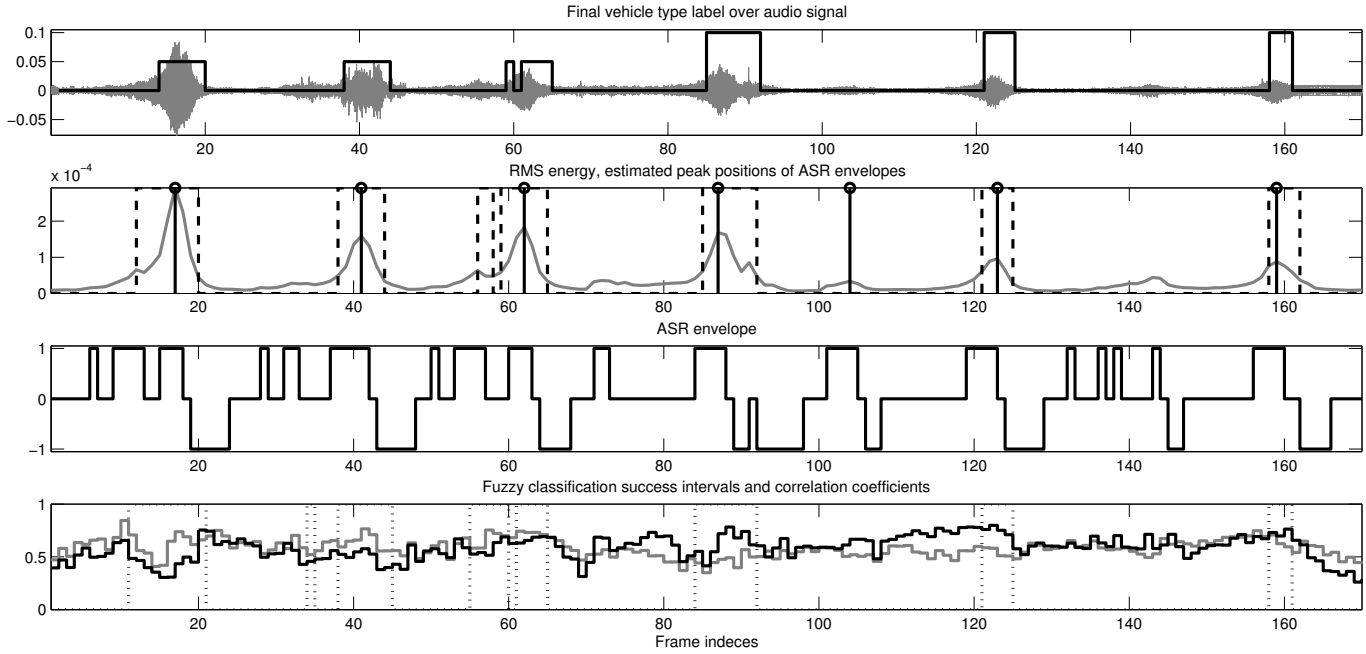

Fig. 7. Algorithm testing results, first test signal. From top to bottom, First Subplot: test signal with 6 instances of passing vehicles (grey); final estimated labels with values 0.05 corresponding to class 1 and 0.1 – to class 2 (black); Second Subplot: RMS energy readings per frame (grey); signal energy threshold (black horizontal line); energy peaks approximated by ASR envelope (black stems); Third Subplot: coded RMS energy dynamic of the ASR envelope; Fourth Subplot: intervals of positive fuzzy membership to vehicle feature subspace (dotted vertical lines); coefficients of correlation to the reference spectral vectors (grey – class 1, black – class 2).

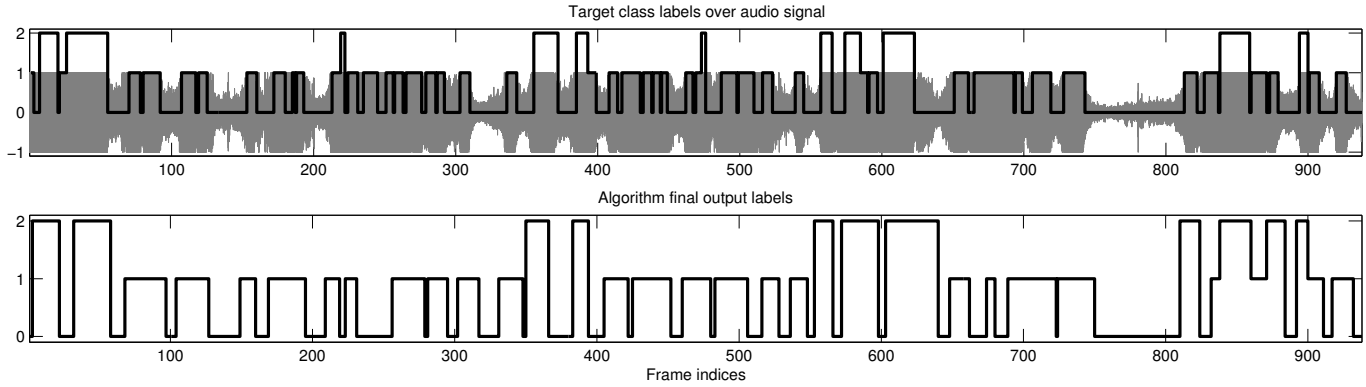

Fig. 8. Algorithm testing results: second test signal. From top to bottom, First Subplot: test signal with instances of passing vehicles (grey), reference labels with values 1 corresponding to class 1 and 2 – to class 2 (black); Second Subplot: final estimated labels with values 1 corresponding to class 1 and 2 – to class 2.

1) Severe pollution of the whole signal frequency band with high levels of ambient noise.
2) Due to dense traffic the time interval between vehicle passes is very short and often does not allow for distinguishing between successive vehicle passes. Furthermore, sounds of vehicles driving on lanes of opposite direction may overlap and distort one another.
3) Sound masking. A heavy truck can emit a noise loud enough to mask the sound of a nearer but lighter car, thus making this car undetectable.
4) Intermediate vehicle types (e.g. minibus or pickup truck) make the boundary between passenger and heavy cars more ambiguous. As a result, for some specific types of vehicles precise classification is not possible.

C. General Testing Results

The algorithm operates sufficiently well in both cases of motor vehicles passing with a certain time interval between the passes and heavy traffic. However if the flow of vehicles is consistent and very dense, a decrease of identification quality is witnessed. The influence of background noise, though reduced due to the algorithm's multistage decision-making logic, cannot be eliminated completely. The algorithm is applicable under different weather conditions, which is demonstrated on the examples of both high and low SNR recordings. Possessing a variety of tunable parameters, the sensitivity of the algorithm can be adjusted to the needed extent. This provides the opportunity to apply the algorithm for classification of various types of moving objects not limited to motorized vehicles.

TABLE I
ALGORITHM OPERATION TIMINGS

Processing times (s)	Algorithm sub-procedures				
	1)	2)	3)	5)	Total
Mean	0.0350	0.0022	0.0110	0.0044	0.0526
Maximum	0.0367	0.0027	0.0119	0.0048	0.0564

VI. REAL-TIME OPERATION ON EMBEDDED DEVICE

For the implementation we choose the embedded devic Gumstix Overo Water with System-on-Chip OMAP3530 (60 MHz ARM-Cortex-A8), 256MB RAM with a 4GB microSI card. The test signal for the real-time operation experiment i similar to the second test signal considered in Section V-B The test signal was acquired prior to the experiment. Afte training the fuzzy rule-base and tuning all the parameters o the algorithm, the signal file is streamed to the device inpu buffer bypassing the ADC at the rate of the sampling frequenc in order to simulate real-time data acquisition and operatio [16]. The frame length is chosen, as in the previous sections to be $2^{14} = 16384$, which corresponds to 0.3715 second at the sampling rate 44.1 kHz. To operate in real-time, th identification procedure therefore must take less than 0.3715 seconds to compute.

The test signal is 625.27 seconds long, which corresponds to 1683 frames of length 16384. The processing time is measured for the following procedures:

1) FFT execution
2) RMS energy calculation
3) Feature extraction
4) Fuzzy Classification
5) Calculation of correlation coefficients
6) ASR envelope estimation

During the experiment the algorithm is made to run to full extent, not terminating during negative detection, i.e. after 2) or 4), in order to achieve consistent results. The mean and maximal values of processing times are presented in Tab. I. Operations 4) and 6) are excluded from the table, as the times of 4) are either 30-31 or rarely 61 μs, and for 6) – 30-31 μs. Thus the mean values do not need to be estimated.

As expected, the most time-consuming operations are FFT execution (consuming more than half of the total processing time) and feature extraction. Process RMS energy calculation takes very little time, so during non-detection the system resources are greatly spared. Correlation computation also consumes much time, growing along with frame length and the number of reference vectors. Thus finding a faster alternative to this method will increase system performance. Altogether, the mean total processing times are significantly smaller than frame durations, thus the algorithm can easily operate in real-time on the given platform.

The distributions of processing times of the computation steps with most variance are presented as histograms in Fig. 9. The variance of processing time is small and thus the predictability of computation time is high. A small amount of values noticeably larger that the mean exist for every sub-procedure and the total process. These abnormalities of long processing durations are most certainly influenced by the

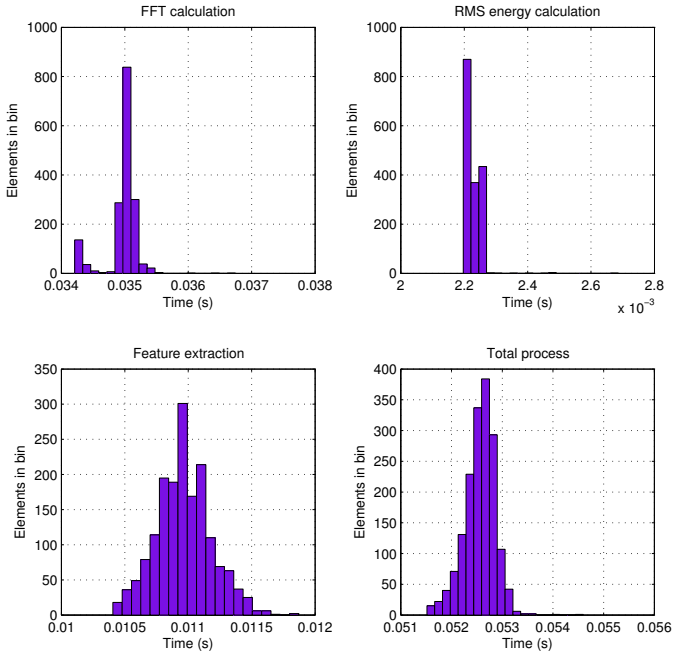

Fig. 9. Histograms of processing times of three sub-procedures with most variance and the whole process.

hardware limitations and problems with memory allocation. Taking this into account and calculating the worst-case total processing time, consisting of maxima of every sub-procedure time, gives 0.0564 seconds for frame length 16384, which is still much less than the signal frame duration.

VII. CONCLUSION

In this paper we have introduced a hierarchical algorithm for mobile vehicle identification by means of acoustic noise analysis. The algorithm is developed specifically for real-time applications and is therefore computationally inexpensive. The hierarchical structure of the algorithm embeds several signal analysis and classification techniques that are applied in a systemized manner to support complex decision-making. The testing results indicate that the algorithm has a potential to detect and classify motor vehicles under varying weather conditions. A possible implementation of the algorithm on an embedded device is presented and its real-time operation capability is experimentally proven.

For future developments the algorithm's robustness may be increased by applying soft discretization to the transitions of the algorithm decision-making path [17] thus transforming its appearance to a fuzzy tree. The final class label therefore may be decided based on degrees of membership. Final class label derivation logic may also be reconsidered to support hierarchy in class label assignment, that will allow for increasing the number of classes by using sub-classes and thus covering the intermediate vehicle types. Another direction of development lies in the integration of other types of sensors in order to enable enhanced environment perception by means of data fusion [3]. On the other hand, application of several microphones in an array configuration would permit to localize the vehicles' positions in the monitored area [18].

REFERENCES

[1] P. E. William and M. W. Hoffman, "Classification of military ground vehicles using time domain harmonics' amplitudes," vol. 60, no. 11, pp. 3720–3731, 2011.

[2] E.-H. Ng, S.-L. Tan, and J. G. Guzman, "Road traffic monitoring using a wireless vehicle sensor network," in *Proc. Int. Symp. Intelligent Signal Processing and Communications Systems ISPACS 2008*, 2009, pp. 1–4.

[3] A. Klausner, A. Tengg, and B. Rinner, "Vehicle classification on multi-sensor smart cameras using feature- and decision-fusion," in *Proc. First ACM/IEEE Int. Conf. Distributed Smart Cameras ICDSC '07*, 2007, pp. 67–74.

[4] T. Takechi, K. Sugimoto, T. Mandono, and H. Sawada, "Automobile identification based on the measurement of car sounds," in *Proc. 30th Annual Conf. of IEEE Industrial Electronics Society IECON 2004*, vol. 2, 2004, pp. 1784–1789.

[5] A. Starzacher and B. Rinner, "Single sensor acoustic feature extraction for embedded realtime vehicle classification," in *Proc. Int Parallel and Distributed Computing, Applications and Technologies Conf*, 2009, pp. 378–383.

[6] N. A. Rahim, M. P. Paulraj, A. H. Adom, and S. Sundararaj, "Moving vehicle noise classification using backpropagation algorithm," in *Proc. 6th Int Signal Processing and Its Applications (CSPA) Colloquium*, 2010, pp. 1–6.

[7] S. Maithani and R. Tyagi, "Noise characterization and classification for background estimation," in *Proc. Int. Conf. Signal Processing, Communications and Networking ICSCN '08*, 2008, pp. 208–213.

[8] S. S. Yang, Y. G. Kim, and H. Choi, "Vehicle identification using wireless sensor networks," in *Proc. IEEE SoutheastCon*, 2007, pp. 41–46.

[9] G. Gritsch, N. Donath, B. Kohn, and M. Litzenberger, "Night-time vehicle classification with an embedded, vision system," in *Proc. 12th Int. IEEE Conf. Intelligent Transportation Systems ITSC '09*, 2009, pp. 1–6.

[10] V. Cevher, R. Chellappa, and J. H. McClellan, "Vehicle speed estimation using acoustic wave patterns," *Trans. Sig. Proc.*, vol. 57, no. 1, pp. 30–47, Jan. 2009.

[11] M. Zivanovic, A. Röbel, and X. Rodet, "Adaptive threshold determination for spectral peak classification," *Comput. Music J.*, vol. 32, no. 2, pp. 57–67, Jun. 2008.

[12] M. Frigo and S. G. Johnson, "FFTw: an adaptive software architecture for the FFT," in *Proc. IEEE Int Acoustics, Speech and Signal Processing Conf*, vol. 3, 1998, pp. 1381–1384.

[13] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the cuidado project," CUIDADO I.S.T. Project Report, Tech. Rep., 2004.

[14] A. Riid and E. Rustern, "An integrated approach for the identification of compact, interpretable and accurate fuzzy rule-based classifiers from data," in *Proc. 15th IEEE Int Intelligent Engineering Systems (INES) Conf*, 2011, pp. 101–107.

[15] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete mathematics: a foundation for computer science*. Addison-Wesley Reading, MA, 1994, vol. 2.

[16] S. Astapov, J. S. Preden, and E. Suurjaak, "A method of real-time mobile vehicle identification by means of acoustic noise analysis implemented on an embedded device," in *Proc. 13th Biennial Baltic Electronics Conf. (BEC)*, 2012, pp. 283–286.

[17] Y. Peng and P. Flach, "Soft discretization to enhance the continuous decision tree induction," in *Integrating Aspects of Data Mining, Decision Support and Meta-Learning*, C. Giraud-Carrier, N. Lavrac, and S. Moyle, Eds. ECML/PKDD'01 workshop notes, September 2001, pp. 109–118.

[18] V. C. Ravindra, Y. Bar-Shalom, and T. Damarla, "Feature-aided localization of ground vehicles using passive acoustic sensor arrays," in *Proc. 12th Int. Conf. Information Fusion FUSION '09*, 2009, pp. 70–77.