Multi-Target Classification Using Acoustic Signatures in Wireless Sensor Networks: A survey

Ahmad Aljaafreh & Ala Al-Fuqaha

Abstract

Classification of ground vehicles based on acoustic signals using wireless sensor networks is a crucial task in many applications such as battlefield surveillance, border monitoring, and traffic control. Different signal processing algorithms and techniques that are used in classification of ground moving vehicles in wireless sensor networks are surveyed in this paper. Feature extraction techniques and classifiers are discussed for single and multiple vehicles based on acoustic signals. This paper divides the corresponding literature into three main areas: feature extraction, classification techniques, and collaboration and information fusion techniques. The open research issues in these areas are also pointed out in this paper. This paper evaluates five different classifiers using two different feature extraction methods. The first one is based on the spectrum analysis and the other one is based on wavelet packet transform.

Keywords: Signal classification, feature extraction, distributed sensors, sensor fusion.

1. INTRODUCTION

Wireless sensor network (WSN) is a network of spatially distributed, densely deployed, and self-organized sensor nodes, where a sensor node is a platform with sensing, computation and communication capabilities. WSN is an emerging technology because of the advances in technologies of: Micro-Electro-Mechanical Systems (MEMS), Microprocessors, wireless communication and power supply. New technologies provide cheap small accurate: sensors, processors, wireless transceivers, and long-life batteries. Sensor node is the integration of all of these technologies in a small board, like the ones in Fig. 3 part (b), it is called mote. Fig. 3 part (a) shows the basic architecture of the mote. All of the above motivate researchers and practitioners to design, deploy and implement networks of these sensor nodes in many applications. WSN has the following characteristics: concern is about the data but not about the sensor node itself, low cost, constrained power supply, static network, topology may change because of sensor node or link failure, sensor nodes are prone to destruction and failure, dense deployment, self-organization, and spatial distribution. WSN is used in many remote sensing and data aggregation applications [1],[2]. Detection, classification, and tracking are the main signal processing functions.
of the wireless sensor networks [3]. WSNs increase the covered area, redundancy of the sensors, and decision makers, which improves the performance and reliability of the decision making. To understand the work, design and operation of the WSNs see Refs. [4],[5]. Refs. [4],[6] categorizes the applications and describes the implementation of the WSNs. A survey of the architecture and sensor nodes deployment in WSNs is presented in Ref. [7]. WSN is a cost efficient technology. However, it has some constraints. Limited energy, limited bandwidth, and limited computational power are the main constraints of WSNs [8]. Therefore, to implement any digital signal processing algorithm it needs to be an intelligent signal processing and decision making algorithm with the following requirements: power efficiency, robustness, and scalability. In WSNs, observed data could be processed at the sensor node itself, distributed over the network, or at the gateway node. WSNs can be utilized for distributed digital signal processing [9]-[11]. Research in classification in wireless sensor networks can be divided into two areas: hardware area (platforms, sensors), and software area (signal processing algorithms, collaboration, and networking techniques) [12]. The signal processing techniques and collaboration schemes that are used in ground vehicle classification in WSN based on acoustic signals are surveyed, as in Fig 2, in this paper. Target classification in WSN is to label or categorize a target that passing through the area that is monitored by the WSN to one of a predefined classes based on an extracted feature vector. Classification in WSNs can be considered as a process as in Fig. 4, where a feature vector is extracted from the input signal, then classified, then the information is fused to come up with the final decision. Most of the researcher are interested in improving the performance of this process through selection and design an efficient tool, as in Table 1, for one of the followings tasks:

- Feature Extraction
- Classification Techniques
- Information Fusion

The remainder of the paper is organized as follows. Section 2 presents the recent methods that are used to extract features from the vehicle acoustic signals for single and multiple targets. Section 3 discusses the classification techniques. Section 4 presents the information fusion techniques. Section 5 outlines the the open research. And finally, conclusions are discussed in section 6.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Feature Extractor</th>
<th>Classifier</th>
<th>Classes Number</th>
<th>Classification Rate</th>
<th>Fusion Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>[12]</td>
<td>TESPAR</td>
<td>ANN</td>
<td>2</td>
<td>up to 100%</td>
<td>-</td>
</tr>
<tr>
<td>[13] and [14]</td>
<td>DWT</td>
<td>MPP</td>
<td>2</td>
<td>98,25%</td>
<td>-</td>
</tr>
<tr>
<td>[15]</td>
<td>HLA, PSD</td>
<td>ANN</td>
<td>4</td>
<td>HLA: 92%, PSD: 94%</td>
<td>-</td>
</tr>
<tr>
<td>[16]</td>
<td>HLA</td>
<td>ANN</td>
<td>18</td>
<td>88%</td>
<td>running sum</td>
</tr>
<tr>
<td>[17]</td>
<td>HLA</td>
<td>MAP</td>
<td>6</td>
<td>89%</td>
<td>-</td>
</tr>
<tr>
<td>[18]</td>
<td>MFCC</td>
<td>GMM, HMM and ML</td>
<td>9</td>
<td>77%, 88%</td>
<td>-</td>
</tr>
<tr>
<td>[19]</td>
<td>FFT, DWT, STFT, PCA</td>
<td>kNN, MPP</td>
<td>4</td>
<td>85%, 88%</td>
<td>MRI</td>
</tr>
<tr>
<td>[20]</td>
<td>STFT, PCA</td>
<td>ANN</td>
<td>3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>[21]</td>
<td>FFT, PSD, AR</td>
<td>kNN, ML, SVM</td>
<td>2</td>
<td>78% - 97%</td>
<td>-</td>
</tr>
<tr>
<td>[22]</td>
<td>DWT</td>
<td>ANN</td>
<td>4</td>
<td>73%</td>
<td>-</td>
</tr>
<tr>
<td>[23]</td>
<td>CC</td>
<td>HMM</td>
<td>9</td>
<td>96%</td>
<td>-</td>
</tr>
<tr>
<td>[24]</td>
<td>WPT</td>
<td>LDA, CART</td>
<td>3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>[25]</td>
<td>CWT</td>
<td>ANN</td>
<td>6</td>
<td>95%</td>
<td>-</td>
</tr>
<tr>
<td>[26]</td>
<td>TVAR, PCA</td>
<td>ANN</td>
<td>6</td>
<td>83%-95%</td>
<td>-</td>
</tr>
<tr>
<td>[27]</td>
<td>BHM</td>
<td>CART</td>
<td>9</td>
<td>90%</td>
<td>Decision Fusion</td>
</tr>
<tr>
<td>[28]</td>
<td>STFT, RID</td>
<td>ANN, MVG</td>
<td>6</td>
<td>up to 87%</td>
<td>-</td>
</tr>
<tr>
<td>[29]</td>
<td>EE, PCA</td>
<td>ANN, Fuzzy</td>
<td>5</td>
<td>up to 97%</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 1: Recent feature extraction and classification techniques used for vehicle classification based on acoustic signals.

<table>
<thead>
<tr>
<th>Feature Extraction Method</th>
<th>Classification Technique</th>
<th>Accuracy</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>[36] Harmonic and Frequency Components</td>
<td>SVM</td>
<td>5</td>
<td>85%</td>
</tr>
<tr>
<td>[37] Harmonic set</td>
<td>MVG</td>
<td>3-5</td>
<td>70-80%</td>
</tr>
<tr>
<td>[38] STFT, PCA</td>
<td>C4.5, KNN, PNN, SVM</td>
<td>4</td>
<td>60-93%</td>
</tr>
<tr>
<td>[39] WPT</td>
<td>CART</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>[40] MFCC</td>
<td>RNN</td>
<td>4</td>
<td>85%</td>
</tr>
<tr>
<td>[41] PSD</td>
<td>KNN, ML, SVM</td>
<td>2</td>
<td>up to 97%</td>
</tr>
<tr>
<td>[42] PSD, PCA</td>
<td>SVM</td>
<td>3</td>
<td>up to 93%</td>
</tr>
<tr>
<td>[43] MFCCs</td>
<td>GMM</td>
<td>2</td>
<td>up to 94.1%</td>
</tr>
<tr>
<td>[44] FFT, WT</td>
<td>KNN, MPP, K-Means</td>
<td>3</td>
<td>95.5%</td>
</tr>
<tr>
<td>[45] WPT</td>
<td>ML, ANN</td>
<td>3</td>
<td>up to 98%</td>
</tr>
<tr>
<td>[46] PSD</td>
<td>ANN</td>
<td>4</td>
<td>up to 99%</td>
</tr>
<tr>
<td>[47] WPT</td>
<td>cascaded fuzzy classifier (CFC)</td>
<td>3</td>
<td>-</td>
</tr>
</tbody>
</table>

2. FEATURE EXTRACTION OF ACOUSTIC SIGNATURE

Feature extraction is the most significant phase of the classification process. To classify an object, a set of features of that object is extracted to label that object to one of a predefined classes. This set of features is generated from a source signal as in Fig. 1. Feature extraction can be considered as dimensionality reduction technique. In feature extraction certain transforms or techniques are used to select and generate the features that represent the characteristic of the source signal. This set of features is called a feature vector. Feature vectors could be generated in time, frequency, or time\frequency domain.
**Figure 1:** Classification block diagram.

**Figure 2:** Taxonomy of the techniques that are used in target classification using acoustic signature in wireless sensor networks.

**Figure 3:** Wireless sensor node examples in part (b) and the common architecture of a sensor node in part (a).
2.1 Time Domain
The computation of feature vector in time domain is usually simple. Ref. [48] discusses two time-domain feature generation methods. The first method is based on the energy distribution of the signal, where the energy of a short time window of the source signal is used to discriminate between classes. The second method is based on counting the number of zero crossings of a signal within a time interval. The energy envelope (EE) in time domain is considered in [29]. Time Encoded Signal Processing and Recognition (TESPAR) is a method that is used in speech waveform encoding. TESPAR is used in [12] to generate features from vehicle acoustic and seismic signals. TESPAR is based on the duration and shape of the portion of the waveform that is between two zero crossings.

Principal Component Analysis (PCA) is popular statistical tools that is used for dimensional reduction. PCA is based on finding the principal eigenvectors of the covariance matrix of the set of signals. PCA is used as a feature extraction method in [19, 20, 38, 42].

2.2 Frequency Domain
Frequency based feature generation methods, like Fast Fourier Transform (FFT), are common approaches in vehicle classification [16], [20], [31], [33]-[35], [49]. In [31] Fast Fourier Transform (FFT) and Power Spectral Density (PSD) are used to extract feature vectors. Similarly in [35], the first 100 of 512 FFT coefficients are averaged by pairs to get a 50-dimensional FFT-based feature vector with resolution of 19.375 Hz and information for frequencies up to 968.75 Hz. Ref. [34] presents schemes to generate low dimension feature vectors based on PSD using an approach that selects the most common frequency bands of PSD in all the training sets for each class. Ref. [33] proposes an algorithm that uses the overall shape of the frequency spectrum to extract the feature vector of each class. Principal component eigenvectors of the covariance matrix of the zero-mean-adjusted samples of spectrum are also used to extract the sound signature as in [20]. Some vehicle acoustic signatures have a pattern of relation between the harmonics amplitude. Harmonics are the peaks of the spectral domain. The relation between the amplitude and the phase of these peaks is used to form the feature vector. Harmonic Line Association (HLA) feature vector is used in [30], where the magnitude of the second through 12th harmonic frequency components are considered as the feature vector to be used for vehicle classification. Different algorithms are used to estimate the fundamental frequency. In [36], two sets of features are extracted from the vehicle sound. The first one is based on the harmonic vector. The second one is a key frequency feature vector. In [37], the number of harmonics is modeled as a function of
the vehicle type. Looking for stable features other than the harmonics relation, Ref. [50] models
the vehicle acoustic signature by a coupled harmonic signal. Cepstral coefficients (CC) are the
coefficients of the inverse Fourier Transform of the log of the magnitude of the spectrum [23].
Mel-frequency cepstral coefficients (MFCC) is used in [18],[40],[43] as a feature extractor, where
the feature vector is made up of few of the lowest cepstrum coefficients. Mel-frequency cepstrum
(MFC) is a representation of the short-term power spectrum of a sound, where the log power
spectrum on a nonlinear mel scale of frequency is transformed based on a linear cosine
transform.

Two types of spectral features are explored in [21]: Non-parametric FFT-based PSD estimates,
and Parametric PSD estimates using autoregressive (AR) modeling of time series. In AR model
the value of any variable at any time is modeled as a function of a finite number of the past values
of that variable. The number of the involved past values is called the model order. AR of the first
order is a Markov chain, where the current value depends only on the previous value. A random
variable \(X\) can be modeled at time \(t\) using AR of order \(P\) as follows:

\[
x_t = \sum_{k=1}^{P} \theta_k x_{t-k} + \omega_t
\]

(1)

where \(\theta_k\) denotes the corresponding autoregressive coefficients. \(\omega_t\) is a white gaussian noise
with zero mean. If \(\theta_k\) is varying with time then the AR process is called Time Varying
Autoregressive (TVAR). TVAR is used to model the acoustic signal in [26, 51]. A filter bank is
used based on the biology based hearing model (BHM) as a feature extraction system in [27].

**Feature Extraction Example using the Spectrum Distribution:**
The goal is to develop a scheme for extracting a low dimension feature vector, which is able to
produce good classification results. The first feature extraction technique of acoustic signals in
this paper is based on the low frequency band of the overall spectrum distribution. The low
frequency band is utilized, because most of the vehicle’s sounds come from the rotating parts,
which rotate and reciprocate in a low frequency, mainly less than 600 Hz as it is clear in Fig. 6.
Sounds of moving ground vehicles are recorded at the nodes at a rate of 4960 Hz as in Fig.5.
After the positive detection decision, a signal of event is preprocessed as the following:

![Figure 5: Time domain for three different sounds for two different vehicles v1 and v1.](image-url)
DC bias should be removed by subtracting the mean from the time series samples.

\[ x_i(n) = x_i(n) - \frac{1}{N} \sum_{n=1}^{N} x_i(n) \]  

(2)

Feature vector will be the median of the magnitude of the STFT of a signal of event. It will be computed as the following: the magnitude of the spectrum is computed by FFT for a hamming window of size 512, without overlapping.

\[ X_i(W) = FFT(x_i(n)) \]  

(3)

After this, the spectrum magnitude is normalized for every frame

\[ X_i(W) = \frac{X_i(W)}{\sum_{W=1}^{K} X_i(W)} \]  

(4)

where \( K \) is the window size. The median of all frames is considered as the extracted feature vector.

\[ X_{\text{f}}(W) = \text{median}(X_i(W)) \]  

(5)

The mean of all frames could also be considered as the extracted feature vector.

\[ X_{\text{f}}(W) = \frac{1}{Z} \sum_{i=1}^{Z} X_i(W) \]  

(6)

where \( Z = N/k \). The first 64 points of the median of the spectrum magnitude contain up to 620 Hz. This gives a 64 dimensional vector that characterizes each vehicle sound. We compared feature extraction using the mean and the median. The median gives better results, specially for noisy environments. Fig.7 displays the acoustic spectral distribution of vehicle 1 and vehicle 2. For the unknown utterance, the same steps are done, except one frame of FFT is considered as the feature to be classified to reduce the computational cost, because this FFT computation is performed online. This can be extended to have multiple frames, but this will increase the computational cost.
2.3 Time Frequency Domain

Short Time Fourier Transform (STFT) is used in [38] to transform the overlapped acoustic Hamming windowed frames to a feature vector. Ref. [52] proposes a probabilistic classifier that is trained on the principal components subspace of the short-time Fourier transform of the acoustic signature. Wavelet transforms provide multi-resolution time-frequency analysis [53]. Wavelet transforms (WT) is the the projection of a signal onto the wavelet. Wavelet is a series of functions \( \psi_{a,b}(t) \) derived from a base function \( \psi(t) \) by translation and dilation.

\[
\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-p}{a}\right)
\]

(7)

where a is called scale parameter, b is called translation or shift parameter, and \( \psi(t) \) is called wavelet base function. Wavelet Transform is called CWT when values of a and b are continuous, and it is called DWT when they are discrete [54]. Discrete Wavelet Transform (DWT) approximation coefficients y are calculated by passing the time series samples x through a low pass filter with impulse response g.

\[
y(n) = x(n) * g(n) = \sum_{k=-\infty}^{\infty} x(k) g(n-k)
\]

The signal is also decomposed simultaneously using a high-pass filter h. The outputs from the high-pass filter are the detail coefficients. The two filters are related to each other. DWT is exactly the same as the Wavelet Packet Transform (WPT) except that in DWT the next level is the result of one step of wavelet transform of the approximation branch and not the detail one. Wavelet packet transform can be viewed as a tree structure. The root of the tree is the time series of the vehicle sound. The next level is the result of one step of wavelet transform. Subsequent levels in the tree are obtained by applying the wavelet transform to the low and high pass filter results of the previous step's wavelet transform. The Branches of the tree are the blocks of coefficients. Each block represents a band of frequency. Feature extraction of acoustic signals is based on the energy distribution of the block coefficients of wavelet packet transform. A wavelet-based acoustic signal analysis of military vehicles is presented in [13, 55]. Discrete Wavelet Transform

Ahmad Aljaafreh & Ala Al-Fuqaha

(DWT) is used in [13] and [14] to extract features using statistical parameters and energy content of the wavelet coefficients. Wavelet Packet Transform (WPT) has a higher frequency resolution than the DWT [56]. WPT is also used to extract vehicle acoustic signatures by obtaining the distribution of the energies among blocks of wavelet packet coefficients like in [24],[39]. Ref. [53] has a proof that wavelet analysis methods is suitable for feature extraction of acoustic signals.

**Feature Extraction Using Wavelet Packet Transform:**

After the positive detection decision, a one second time series is preprocessed as the following:

- The wavelet packet transform is applied for this signal then the energy of each block coefficients of the (L) level is calculated.
- This approach provides a vector of length = original time series length /2^L . Which is considered the feature vector.

![Wavelet block energy distribution](image)

**Figure 8:** Wavelet block energy distribution for vehicle one in first row for three different sounds and for vehicle 2 in the second row.

Fig. 8 displays the blocks energy distribution for vehicle 1 and vehicle 2. In this paper we used classification rate as the metric for the evaluation of the feature extraction performance. But this metric depends on the classifier itself. Thus, we compare the classification rate for two classifiers as shown in Fig. 12.

### 2.4 Feature Extraction Performance Using Separability Measures

Separability measures provide a measure of class discriminability based on feature space partitioning. A good feature vector extractor provides close feature vectors for the same class, and far feature vectors for distinct classes. The goal is to have a feature extraction method that has high distance between distinct classes and low distance within each class. The metric, in this paper, is the separability ratio (sr), which is the ratio between the intraclass distance and the average interclass distance [57].

\[
\text{sr} = \frac{D_g}{D_i}
\]

\[
D_g = \sum_{i=1}^{C} \frac{P}{n_i} \sum_{k=1}^{n_i} ((V_{ik} - m_i)(V_{ik} - m_i)^T)^{1/2}
\]

![Equation 8 and 9]

(8)  

(9)
$D_g$ represents the average of the variances of distance within all classes. $V_{ik}$ is the normalized feature vector. $C$ is the number of classes. $P_i$ is the probability of class $i$. $n_i$ number of vectors in class $i$. $m_i$ is the mean vector for class $i$.

$$D_g = \frac{1}{C} \sum_{i=1}^{C} P_i \left( \frac{1}{n_i} \sum_{k=1}^{n_i} (m_{ik} - m)(m_{ik} - m)^T \right)$$

$D_l$ represents the average of the distances between all classes. $m$ is the mean for all classes.

$$m_i = \frac{1}{n_i} \sum_{k=1}^{n_i} V_{ik}$$

$$m = \frac{1}{n} \sum_{i=1}^{C} \sum_{k=1}^{n_i} V_{ik}$$

The smaller the ratio is the better the separability. This means that the best feature extraction scheme is the one that decreases $D_g$ and increases $D_l$.

### 3. Classification Techniques

Classifiers provide the functions or the rules that divide the feature space into regions, where each region corresponds to a certain class. Having a number of $N$-dimensional feature vectors for each class, a function can be deduced to partition the $N$ feature space to number of regions, where each region represents a class. This process is called supervised learning or classification. Classifiers can be categorized to parametric or non-parametric. Some researches combine multiple different classifiers, called compound classifiers.

#### 3.1 Parametric Classifiers

Parametric classifiers are the classifiers that can be represented in closed-form. For instance, assuming that the distribution of a certain class as a parametric form such as Gaussian. Some classifiers are based on discrimination function with a certain parametric form such as support vector machine. Below are the most parametric classifiers that have been used in vehicle classification based on acoustic signature.

##### 3.1.1 Bayesian Classifier

Bayesian classifier is a probabilistic classifier based on using Bayes’ theorem. Maximum likelihood (ML) is used to estimate the Bayesian classifier parameters. Maximum A posteriori Probability (MAP) can also be considered as a generalization of ML. Each class is assumed to be independent instances of parametric distributed random process. A naive Bayes classifier is a variation of the Bayesian classifier with assumption of an independent feature model. Bayesian classifier is used in many research papers with assumption that each class is a normal distributed random process [35],[41],[58],[59] where features vector $S$ of each class $C$ is assumed to be independent instances of normally distributed random process.

$$p(S | \theta_i) = N(\mu_i, \Sigma_i)$$
\( \mu_i \) and \( \Sigma_i \) are the mean and covariance matrix respectively. \( \theta_i \) is the parameter set of \( i^{th} \) distribution, \( \theta_i = \{ \mu_i, \Sigma_i \} \).

\[
p(S | \theta_i) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} (S - \mu)^T \Sigma^{-1} (S - \mu)\right]
\]

To represent each training set of each class as a distribution with \( \hat{\Sigma} \) and \( \hat{\mu} \) parameters the likelihood of \( \hat{\theta} \)

\[
l_i(\theta) = \sum_{k=1}^{n} \ln P(s_i | \theta)
\]

should be maximized by equating \( \Delta l_i = 0 \), then ML estimations of \( \mu \) and \( \Sigma \) are

\[
\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} s_k
\]

\[
\hat{\Sigma} = \frac{1}{n} \sum_{k=1}^{n} (s_k - \hat{\mu})(s_k - \hat{\mu})^T
\]

for minimum error classification the a posteriori probability should be maximized

\[
p(\theta_i | S) = \frac{p(S | \theta_i) p(\theta_i)}{\sum_{j=1}^{C} p(S | \theta_j) p(\theta_j)}
\]

\( h_i(x) \) denote the logarithmic version of \( p(\theta_i | S) \)

\[
h_i(x) = \ln p(S | \theta_i) + \ln p(\theta_i) - G
\]

where \( G \) is constant can be ignored in the optimization

\[
h_i(x) = -\frac{1}{2} (S - \mu)^T \Sigma^{-1} (S - \mu) - D
\]

where \( D \) is constant then any vehicle feature vector \( x \) is classified to class \( i \) according to the discriminant functions \( h_i(x) \) if \( h_i(x) > h_j(x) \) for all \( j \neq i \). Linear discriminant Analysis (LDA) assumes that the class covariances \( \Sigma_i \) are identical. LDA is used as a linear classifier in [24] ML is the optimum classifier but it needs large number of training set. Training ML classifier with small number of training set will not give an invertible covariance matrix. This makes it hard to compute the discriminant functions.
3.1.2 Support Vector Machine (SVM)

SVM is widely used as a learning algorithm for classifications and regressions. SVM classify data \( x_i \) by class label \( y_i \in \{+1,-1\} \) given a set of examples \( \{x_i, y_i\} \) by finding a hyperplane \( wx + b \), \( x \in \mathbb{R}^n \) which separate the data point \( x_i \) of each class. as in Fig. 9.

\[
g(x) = \text{sign}(wx + b)
\]

where \( w \) is the weight vector, \( b \) is the bias. SVM choose the hyperplane that maximize the distance between the hyperplane and the closest points in each feature space region which are called support vectors. So the unique optimal hyperplane is the plane that maximize this distance

\[
\frac{|wx_i + b|}{|P_w|P}
\]

This is equivalent to the following optimization problem

\[
\min_{w,b} \frac{P_w P^2}{2}, \text{ s.t. } y_i (w^T x_i + b) \geq 1
\]

For the cases that nonlinear separable, a kernel function maps the input vectors to a higher dimension space in which a linear hyperplane can be used to separate inputs. So the classification decision function becomes:

\[
\text{sign}(\sum_{i \in \text{SVs}} \alpha_i^0 y_i K(p, p_i) + b)
\]

where SVs are the support vectors. \( \alpha_i^0 \) and b are a lagrangian expression parameters. \( K(p, p_i) \) is the kernel function. It is required to represent data as a vector of a real number to use SVM to classify moving ground vehicles. Performance of SVM classifier for vehicle acoustic signature classification for both feature extraction methods is also evaluated. It is found that SVM is a good classifier for stochastic signal in WSN [12], [42], [60]–[63].
3.1.3 Gaussian Mixture Model (GMM)
Due to the constraints in WSN resources, parametric models such as Gaussian mixture model is preferred to non-parametric models [43]. Modeling of acoustic signal in WSN using a parametric model, like GMM requires little resources, and has a good pattern matching performance [43]. GMM is a statistical method that is used for classification and clustering. GMM is a linear combination of M-Gaussian pdfs. Let \( x \) be a N-dimensional feature vector, then the distribution of \( x \) is as follows:

\[
f_m(x) = \sum_{i=1}^{m} \alpha_i \phi(x; \theta_i)
\]

where \( \sum_{i=1}^{m} \alpha_i = 1 \), \( \alpha_i \geq 0 \), \( i = 1, \ldots, m \)

\( \alpha_i \) is the mixing weight, \( \phi(x; \theta_i) \) is the Gaussian mixture component. Component \( i \) has \( N \)-variate Gaussian density function with weight \( \alpha_i \), mean vector \( \mu_i \), and covariance matrix \( \Sigma_i \).

Expectation maximization (EM) is one of the common algorithm that is used to obtain the GMM parameters \( \Phi_i = (\alpha_i, \mu_i, \Sigma_i) \) from the training set. The GMM generated from the training set will be used in vehicle classification as in Fig. 10.

![Main block diagram of pattern recognition](image)

Figure 10: Main block diagram of pattern recognition

Any vehicle feature vector \( x \) is classified to class \( C_i \) if it maximizes \( p(C_i | x) = p(x | C_i) p(C_i) \). If all the classes are assumed to occur with the same probability, then the concern is to maximize \( p(x | C_i) \) for every possible class. GMM is used as a classifier in WSN based on the features that are extracted from the vehicle sounds in [43]. Ref. [43] concludes that the GMM, as a parametric classifier, outperforms the KNN and SVM classifiers, and it also concludes that GMM needs relatively less resources.

3.1.4 Hidden Markov Model (HMM)
Acoustic signals could be modeled as HMM. HMM has a specific discrete number of unobserved states; each state has a transition probability to any other state and an initial probability. Each state may be considered as representing a certain sound of the vehicle [23]. Ref. [23] models the cepstral coefficients that are obtained from the time domain signal as HMM, where the pdfs of the states are assumed to be Gaussian with non-zero means and with a diagonal covariance matrix. Modeling the vehicle sounds as HMM is based on the assumption that the acoustic signal of the vehicle is consisting of a sequence of a discrete number of sounds, where the statics of each sound of these sounds is described by a separate state. The parameters of the HMM are: the state transition probability to any other state, the initial probability for each state, and the
observation pdf parameters for each state. Estimation of the maximum likelihood parameters of the HMM given a data set of the vehicle sounds can be done by a special case of the Expectation-maximization algorithm called the Baum-Welch algorithm; it is also known as the forward-backward algorithm. HMM implementation for vehicle classification is based on the estimation of the sequence of states, given a sequence of observations. Some known algorithms are used for that such as the Viterbi algorithm. GMM is static pattern model, while HMM is a sequential pattern model.

3.1.5 Minimum Distance Approach
Minimum Distance Approach (MPP) is a simple parametric classifier that is based on the minimum distance between the feature vector and the average of the class distribution. MPP assume the distribution of the training set of each class as Gaussian distribution. MPP measure the distance between the average of each class distribution and the feature vector of the test data, then it corresponds the test data to the class that has the minimum distance. MPP is used as a classifier in Refs [13],[19],[44].

3.2 Non-parametric Classifiers
In this kind of classifiers no assumptions are made about the probability density function for each class.

3.2.1 KNN classifier
KNN is a simple and accurate method for classifying objects based on the majority of the closest training examples in the feature space. KNN is rarely used in wireless sensor networks because it needs large memory and high computation. In our experiments we set K to be three. So the KNN classifier finds the closest three neighbors out of all the training set. After that, the KNN classifier counts how many of these three is in class one and how many is in class two then based on the majority classify the tested one. KNN is implemented in many literatures as a benchmark to evaluate other classifiers [21],[31],[34],[35],[44].

3.2.2 Artificial Neural Network (ANN)
Artificial neural networks are a learning intelligent tools used to solve problem that hard to be modeled analytically. A key feature of neural networks is the ability to learn from examples of input/ output pairs in a supervised fashion. In [40], rough neural network is used to recognize type of vehicles in WSN. The network has 25 input neurons, corresponding to the 25-dimisional feature vector, 25 hidden layer neurons, and 4 output neurons. Neural network classifier and the maximum likelihood classifier are compared concerning their advantages and disadvantages in [45]. Artificial neural networks were used as a technique to classify and identify vehicles based on acoustic signals in [46], where a two-layer back propagation neural network is trained by the output of a self organized maps neural network.

3.2.3 Fuzzy Logic Rule-Based classifier
Fuzzy Logic Rule-Based classifier (FLRBC) maps the input to the output based on some rules [64]. Fuzzy logic inference is a simple approach to solving a problem rather than attempting to model it mathematically. Empirically, the fuzzy logic inference depends on human's experience more than the technical understanding of the problem. As in Fig. 11, fuzzy logic inference consists of three stages:
1. Fuzzification: map any input to a degree of membership in one or more membership functions, the input variable is evaluated in term of the linguistic condition.
2. Fuzzy inference: fuzzy inference is the calculation of the fuzzy output.
3. Defuzzification: defuzzification is to convert the fuzzy output to a crisp output.
Cascaded fuzzy classifier is proposed in [47] to classify ground moving vehicles locally at sensor nodes in WSN.

### 3.2.4 Decision Tree

Decision tree is a nonlinear classifier that depends on a multistage decision system, where the classes are sequentially rejected until reaching the accepted class. This kind of classifier splits the feature space into unique regions, where each region represents a class [65]. In Refs. [38],[43],[66] decision tree is utilized as a classifier. C4.5 algorithm is used to generate a decision tree in [38]. Decision tree is sometimes described as classification tree or regression tree. Classification And Regression Tree (CART) analysis is used to refer to both of classification and regression.

### 3.3 Classifiers Evaluation and Comparison

Classification evaluation is to measure how accurate the classifier is. There are three main classification metrics that are used to evaluate the performance of the classifier namely: Correct classification rate, detection probability, and false alarm probability. Correct classification rate is the ratio of the correctly classified events from all the samples for all classes. Detection probability for a certain class is the ratio of the correctly classified events from the samples of that class. False alarm probability for a certain class is the ratio of the wrongly classified events for that class from all samples of other classes. Classifiers might have a high accuracy if tested with the same training data. Thus, a classification validation is needed. There are two common methods that are used for classification validation. The first is called hold-out or split-sample method, where a single part of the data is set aside for testing. The second method is the cross-validation method. k-fold cross-validation divides the data into k subsets. The classifier is trained k times, each time leaving out one of the subsets from training to be used for testing the classifier. If k equals the sample size, the method is called leave-one-out. Ref [45] compared the recognition rate and the robustness of two classifiers, neural network classifier and maximum likelihood classifier. Neural fuzzy techniques for classification of vehicle acoustic signal is used in Ref. [29]. In [38], Four different classifiers decision tree (C4.5), KNN, probabilistic neural network (PNN) and SVM are compared. The classification results indicate the performance of SVM outperforms C4.5, KNN, and PNN. SVM, ML, and KNN are used in [35] to evaluate their data set. In this paper, five different classifiers are compared as in Fig. 12.
Figure 12: A comparison of the correct classification rates for different classifiers for two different feature extraction methods. The first one is base on the spectrum analysis and the other one is based on wavelet packet transform.

3.4 Single Target Classification
Single vehicle classification is to label or classify a vehicle to one of the predefined classes. Classification is based on the feature vector, which is generated from the observed continuous stochastic signal. This stochastic signal could be acoustic, seismic, electromagnetic, or any other kind of signals. The feature vector will be the input of the classifier. The classifier could be any kind of the classifiers that are mentioned in the classifiers section. Classifiers could be parametric or non parametric. The parameters of the parametric one are estimated from a training set. Non parametric classifiers are also trained by a training set. Single vehicle classification techniques can be used for multiple vehicle classification after sources separation or with the assumption that the sensor will not observe two or more vehicles at the same time. This assumption is not realistic, especially in battlefield scenarios.

3.5 Multiple Targets Classification
Many researchers assume that multiple targets could be separated in time and space. However, in many of the surveillance applications this assumption is not realistic, where multiple targets can exist in the sensing range of one sensor at the same time as in Fig. 13. This makes the sensor observes a combination of signals. The combination of multiple signals can be modeled as linear, nonlinear, or convoluting combination of the single target signals. Ref. [67] exploits the classifier that is trained in single target classification to classify a convoy of vehicles. Most of the literature models the multiple targets classification problem as a Blind Source Separation (BSS) problem. BSS problem has been tackled in the literature in many ways, such as neural network [68]-[70] and Independent Component Analysis (ICA). ICA is frequently used to separate or extract the source signals [71]-[75]. In [76] the source extraction problem in wireless sensor network is studied in two different sensor network models. Fast fixed-point ICA algorithm is used for source separation [51] presents a statistical method based on particle filtering for the multiple vehicle acoustic signals separation problem in wireless sensor networks. In [77], a recognition system links BSS algorithm with an acoustic signature recognizer based on missing feature theory (MFT). The result of the comparison between FasICA, Robust SOBI, and their work shows that both of former algorithms are better for mixtures of two signals and more. Refs. [78],[79] discuss problem of source estimation in sensor network for multiple targets detection, then a distributed source estimation algorithm is developed. These solutions have some drawbacks that make it hard to be implemented in WSNs. It is evident that the manager nodes need to perform source separation and source number estimation. These tasks are computationally intensive when executed on the individual sensor nodes. The manager node does the following: estimation of the number of sources, separation or extraction of sources, classification of sources. [80] presents a system that is able to recover speech signal in the presence of additive non-stationary noise. This done
through a combination of the classification and mask estimation. Ref. [37] uses a multi-variate Gaussian classifier for classifying individual acoustic targets after beamforming the received signal in the direction of the targets. We direct the reader for more information in beamforming to [81]. Classification of multiple targets without the separation of the sources based on multiple hypothesis testing is an efficient way of classification [82]. A distributed classifiers based on modeling each target as a zero mean stationary Gaussian random process and the same for the mixed signals is proposed in Ref. [83].

4. COLLABORATIVE CLASSIFICATION

Data fusion, information fusion, data aggregation, multisensor integration and sensor fusion are terms that have been used interchangeably in the literature to refer to the idea and theory of fusion. In WSNs data fusion is needed for the following reasons: sensor failure, sensor and technology imprecision, limitations in spatial and temporal coverage [84]. Information fusion can be classified according to the relationships between the sources as: redundant, cooperative, or complementary. It also can be classified according to the abstraction level of the processed data as: low-level (raw data), medium-level, high-level (decision fusion), or multi-level fusion.

The main objective of collaboration classification is to extract the most beneficial information from the collected data. Because every target has its own signature according to the type of generated signal. Deployment of different kinds of sensors, in the same sensor node or in different sensor nodes, increases the performance of collaborative signal processing. This stems from the fact that every sensor type has a different interference and a measurements noise. Efficient and reliable decision making needs data fusion and collaborative signal processing. Distributed classification algorithms fuse signal or decisions from multiple sensor nodes, then classify the targets based on a priori statistical information [85],[86]. Collaboration could be across the sensor nodes, or within a sensor node only when it include multiple modalities of data. Ref. [67] shows the improvement in classification error because of the collaboration. WSNs have two kinds of collaboration signal processing models: data fusion and decision fusion. For more information in data and decision fusion see [87],[88]. Ref. [89] analyzes both models. Exploiting the collaboration among the sensor nodes enhances even the sensing coverage for the network [90]. Decision fusion has less accuracy than data fusion. However, data fusion has more computation and communication overhead than the data fusion. In decision Fusion, Sensor nodes do some processing then send the decision to the manager node as in Fig. 15, where these decisions could be hard or soft decisions [91]-[93]. Manager node fuses all the decisions and come up with the best decision. Rules of decision fusion could be based on: voting, averaging, Bayesian, Dempster-Shafer (DS) sum or other rules as in [94]. An example of decision fusion is the tracking system that is proposed in [95], where detection and classification are performed in the sensor
node while tracking is performed in the sink node. Data and decision fusion are increasingly implemented in sensor network because of hardware improvement, advances in processing methods, and advances in sensor technology [96]. Fig 14 shows some of the data fusion applications. Data and decision fusion techniques answer the question of how to combine the data and decisions from the sensor nodes in a network to obtain a final decision with optimal resource consumption. Sensor nodes make the measurements, then send raw measurements, processed measurements, or local posterior to the fusion center. Fusion architecture can be hierarchical or distributed. In hierarchical fusion, the data or decision is fused from the sensor node to the higher level fusion center. While in distributed fusion architecture, the data or decision is broadcasted to all other fusion centers.

There are various scenarios of data and decision fusion for single and multiple sensors within the sensor node or cross over the network. Ref. [97] has a survey that focuses on the decision fusion. Ref. [98] studies a distributed decision fusion, where the local decision makers send the ranking of the hypotheses, rather than just the most likely hypothesis. A consensus algorithm which weighs the measurements of the neighboring nodes is presented in [99].

Data fusion from seismic and acoustic signal improves the classification accuracy [67]. In Ref. [43], a decision tree generated by the classification and regression tree algorithm is used to fuse the information from heterogeneous sensors. Multi-resolution integration (MRI) algorithm is used in [19] as a data fusion algorithm to handle the uncertainty of sensor output. MRI algorithm is based on building an overlap function from the outputs of the sensors, where this function is resolved at finer scales.

5. OPEN RESEARCH

For single node processing, researcher use mathematical or statistical tools to extract the features that can represent a unique signature for each vehicle or class. Then another tools are used to classify any new utterance to one of the predefined classes. The main goals of this kind of research is either to increase the performance of the classification or decrease the complexity. The relation between performance and complexity is a trade-off relation. Thus the research is open in both areas. Optimal classifier is the classifier that increase the performance and decrease the complexity. Thus, it will be so beneficial to have a standard metrics that have both measures and have a public data base where researcher can evaluate their algorithms based on these metrics using a public data base. Single node processing versus collaborative processing is a hot area of research. Collaborative processing answers the question of how to combine the data and decisions from the sensor nodes in a network to obtain a final decision with optimal resource consumption. Collaborative processing is an open area for research. Fusion modeling, where the signal to noise ratio for both acoustic and communication channel are considered, is critical to find the optimal number of decision makers. Utilization of time and spatial correlation across different nodes is crucial in collaborative processing. All of the above is related to the signal processing techniques that are used for feature extraction, classification, and data and decision fusion. However this is related to the communication protocols that are used in WSN.
6. CONCLUSIONS
The recent research related to classification process of ground vehicles in wireless sensor network, based on acoustic signals, is reviewed in this paper. Classification process involves three main components: feature extraction, classifier design and selection, and information fusion. Feature extraction methods are classified based on the extraction domain to time, frequency, and time/frequency. Classifiers are also categorized into parametric and non parametric classifiers. In wireless sensor networks decision fusion is preferred rather than data fusion because of the power constraints. This paper proposed two feature extraction methods. One is based on wavelet packet transform and the other is based on spectrum distribution. both methods give almost the same separability and classification rate.

Figure 14: Data fusion applications.

Figure 15: One cluster of a wireless sensor network.
3. REFERENCES


[29] S. Somkiat “Neural fuzzy techniques in vehicle acoustic signal classification”. Ph.D.
dissertation, chair-Vanlandingham, Hugh F. 1997


[57] Wang, Lipo, Fu, and Xiuju. “Data Mining with Computational Intelligence”. Springer Berlin Heidelberg, 2005


Middleware, First International Conference, 2006


[66] B. Lantow. “Applying distributed classification algorithms to wireless sensor networks a brief view into the application of the sprint algorithm family”. In Networking, ICN, Seventh International Conference, 2008


[77] H. Qi, X. Tao, and L. H. Tao. “Multiple target recognition based on blind source separation and missing feature theory”. In Computational Advances in Multi-Sensor Adaptive Processing 1st IEEE International Workshop on Volume, 2005


[93] A. Aljaafreh and L. Dong. “Cooperative detection of moving targets in wireless sensor network based on fuzzy dynamic weighted majority voting decision fusion”. In Networking, Sensing and Control (ICNSC), International Conference, 2010


