Acoustic Information Fusion for Ground Vehicles Classification

Baofeng Guo, M. S. Nixon
School of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, UK
Email: {bg|msn}@ecs.soton.ac.uk

T. Raju Damarla
U.S. Army Research Laboratory
2800 Powdermill Road, Adelphi, MD 20783-1197
Email: rdamarla@arl.army.mil

***Abstract*—Many acoustic factors can contribute to the classification accuracy of ground vehicles. Classification based on a single feature set may lose some useful information. To obtain more complete knowledge regarding vehicles' acoustic characteristics, we propose a fusion approach to combine two sets of features, in which various aspects of acoustic signatures are emphasized individually. The first set of features consists of a number of harmonic components, mainly characterizing engine's noise. The second set of features is a group of key frequency components, designated to reflect other minor but also important acoustic factors, such as tires' friction noise. Fusing these two sets of features provides a more complete description of vehicles' acoustic signatures, and reduces the limitation of relying one particular feature set. Further to a feature level fusion method, we propose a modified Bayesian based fusion method to take advantage of matching each specific feature set with its favored classifier. To assess the proposed fusion method, experiments are carried out based on a multi-category vehicles acoustic data set. Results indicate that the fusion approach can effectively increase the classification accuracy compared to those using each individual set of features. The Bayesian based decision level fusion is found to be significantly better than the feature level fusion approach.**

Keywords: Acoustic vehicle classification, information fusion, feature extraction, mutual information, Bayesian decision fusion.

I. Introduction

Acoustic sensors can collect acoustical signals to identify the type of running ground vehicles. Acoustic sensors can be used in many sensors networks for the applications such as battlefield monitoring and surveillance. They become more and more attractive because of their rapid deployability and low cost [1]–[4]. In acoustic sensors processing, classification algorithms play a critical role to identify the type of vehicle, and help to improve the performance of tracking [3], [5].

Many acoustic features can be extracted for classification of running vehicles. The commonly-used features are the levels of various harmonics [6], [7]. The harmonics features have achieved good classification performance, with a stable and compact representation [3], [5]. Although many encouraging results on acoustic vehicle classification have been shown in the previous research [1], [3]–[5], it still remains a challenging problem due to the complexity of vehicle acoustic signals, the great variation of ambient interferences, etc.

In particular, most classification algorithms that have been developed for acoustic vehicles classification only consider one major feature set. However, the overall acoustic signal of a running vehicle could be much more complicated; the vehicle's sound may come from multiple sources, not exclusively by the engine, but also from tires, brakes, etc. Relying on one particular feature extraction approach is therefore likely to loss information. This could become even worse when the number of models' parameters is further restricted by other factors, such as the dimensionality of a classifier' input.

In this paper, we focus on information fusion approaches for acoustic vehicle classification. We argue that the capability gap between different feature sets can provide potential to improve the classification accuracy by information fusion. Moreover, the information fusion may alleviate the constraint on the input's dimensionality for certain classifiers. For example in a decision level fusion, several classifiers can be applied to each feature set individually, and the overall dimensionality of input is divided by the number of the classifiers.

In the proposed fusion approach, a group of new features are firstly extracted to amend the existing harmonic features. The added features are named as key frequency components, and they are selected by mutual information (MI), a metric based on the statistical dependence between two random variables [8], [9]. Selection of the key acoustic features by the mutual information can help to retain those frequency components that contribute most to the discriminatory information, meeting our goal of fusing information for classification.

To keep the same dimensionality as the original feature space, a feature level fusion is first designed by replacing the higher order (or other less important) harmonic components with the same number of key frequency components. For the purpose of fusion, the key frequency components are deliberately selected to be unrelated with the fundamental frequency. This scheme adds no extra cost in the classification algorithm, but has potential to increase discriminatory capability. Next, an improved Bayesian based decision level fusion is proposed to take advantage of matching each specific feature set with its preferred classifiers. To assess the proposed MI-based acoustic feature extraction and the subsequent fusion methods, experiments are carried out based on a multi-category vehicles acoustic data set.

The rest of this paper is organized as follows. In Section II, we argue that multiple feature sets are needed to improve the vehicles' classification accuracy. Next in Section III, we discuss how to use the mutual information to extract the key frequency components to obtain the necessary new information. Subsequently, to combine the harmonics features and the key frequency features, we design a feature level fusion in Section IV-A and propose a modified decision level fusion in Section IV-B. Experimental results are presented in Section V. Finally, we end this paper with conclusion in Section VI.

II. Using multiple feature sets for acoustic vehicle classification

Differing from the previous research [1], [3]–[5], we first argue that multiple feature sets should be considered for a more effective acoustic vehicle classification.

It is known that the acoustic signature of a running vehicle is made up of a number of individual elements, such as engine noise, tire friction noise, etc. Many classification algorithms that have been developed in acoustic vehicle classification were based on the harmonic features, and have been shown effectiveness [1], [3]–[7]. However, our further discussions can suggest that the harmonics features may be incapable to capture the whole acoustic signatures. For examples, the tire noise is generated by the friction between the tires and road. The useful information embedded in this noise may not necessarily relate with the fundamental frequency and its integral multipliers. This indicates that the harmonics may be unable to capture the useful distinguishing information in this particular element.

Though the tire friction noise seems to be a minor constituent of the whole vehicle's sound, it could contain valuable acoustic signature, and sometimes could be important to vehicle classification. For example, the tires friction noise can reflect the information regarding tires' thread and rubber blocks. These factors are closely linked with the type of vehicle, and should not be omitted for classification. Therefore, to improve the accuracy of acoustic vehicle classification, we propose to apply information fusion to include more useful acoustic information.

In the proposed fusion approach, two sets of features are extracted individually to capture different aspects of the acoustic signature. The first one is a commonly-used harmonic feature vector [3], [5]–[7], named as $\mathbf{x}_h$, which is used to account for the engine noise. The second one is a key frequency feature vector, named as $\mathbf{x}_k$, which is aimed at other useful information, such as the acoustic signature embedded in the tires' friction noise.

Based on the above feature extractions, the amended acoustic signature consists of two parts, $\mathbf{x}_h$ and $\mathbf{x}_k$, respectively. To explore this structure, a natural approach is by data fusion [10]. Because two sets of features characterize the acoustic signals from different aspects, combining them has potential to provide more information regarding the the desired vehicle acoustic signature.

The methods on extracting the harmonic features $\mathbf{x}_h$ can be found in [6], [7]. Thus, the major problems remained in this fusion approach are:

- How to select the key frequency features $\mathbf{x}_k$, which will be discussed in Section III; and
- How to develop a suitable fusion scheme, which will be discussed in Section IV.

III. Extracting new features for harmonics-based vehicle classification

According to our discussions in Section II, the feature vector $\mathbf{x}_k$ is intended to provide different information to the harmonics feature vector $\mathbf{x}_h$. Thus, a practical solution to extract $\mathbf{x}_h$ is by searching the residual inharmonics for a group of key frequency components, in which the contained information will be naturally differ from the harmonics.

To find the key frequency components, an ideal search metric would be the classification accuracy or inversely the Bayes classification error. However, feature selection by directly minimizing the Bayes error is difficult to be analytically performed, and an alternative discriminatory metric has to be sought. In this research, we applied an effective feature selection method based on mutual information, which has been developed in our previous project [9], [11], [12]. To keep completeness of presentation, we first recapitulate part of the important equations as already described in [9], [11].

A. Feature selection by mutual information criterion

Mutual information measures the statistical dependence between two random variables and so can be used to evaluate the relative utility of each feature component to classification [8], [9]. Considering that the Bayes error is bounded by mutual information [13], the key frequency components selected by mutual information analysis is actually approximated to a criterion by optimizing the Bayes error. Therefore, a reliable and realistic performance can be assured to some extent (i.e., in the meaning of Bayes error bound). Also, the implementation of mutual information needs relatively lower computational cost [14], [15], which further makes it more attractive than other metrics.

In information theory, the mutual information is a quantity that measures the mutual dependence of the two variables, and is defined as:

$$I(X,Y) = \int_Y \int_X p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}\, dx\, dy, \qquad (1)$$

where $p(x,y)$ is the joint probability density function of continual random variables X and Y, and $p(x)$ and $p(y)$ are the marginal probability density functions respectively. Mutual information is related to entropy as:

$$\begin{aligned} I(X,Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y), \end{aligned} \qquad (2)$$

given the Shannon entropy (discrete) defined as:

$$H(X) = -\sum_X p(x) \log p(x). \qquad (3)$$

According to the definition of mutual information (1) and its relations to the entropy (3), the use of mutual information for key frequency selection can be initially justified as follows.

Let Y be a random variable standing for the class label (e.g., the vehicle type), and X be another random variable denoting the amplitude for a frequency bin. The entropy $H(Y)$ is known to be a measure of the amount of uncertainty about Y (i.e., the objective of prediction), while $H(Y|X)$ is the amount of uncertainty left in Y when knowing an observation X. From (3), $I(X,Y)$ is the reduction in the uncertainty of class label Y by the knowledge or measurement obtained at frequency bin X. Hence, mutual information can be interpreted as the amount of information that the feature at frequency bin X contains about the class label Y (see Venn diagram in Figure 1). In other words, mutual information is capable to reflect the amount of information that a frequency bin X contains about the class label Y. Since the variable defined by class label is the required classification result, the mutual information measures the capability of using this frequency bin to predict the class label, i.e., the vehicle's identity.

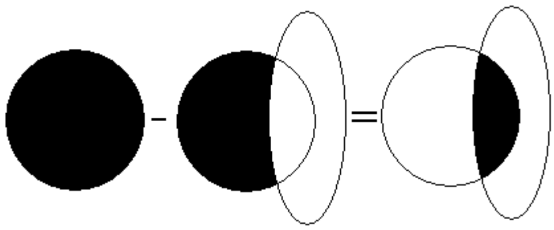

Figure 1. Illustration of mutual information

B. Maximization of mutual information

After the above justifications, we show how to select the key frequency components based on mutual information. The framework of the MI-based feature selection can be described as follows: given a set of original feature vectors $\mathbf{x}'$ with M components or variables, and Y the corresponding output class label (e.g., the vehicle type), find a subset variables $\mathbf{x} \subset \mathbf{x}'$ with N components ($N < M$) that maximizes MI $I(\mathbf{x}, Y)$, i.e.,

$$J\left(\mathbf{x}^0\right) = \max_{\mathbf{x}\subset\mathbf{x}'} I\left(\mathbf{x}, Y\right). \tag{4}$$

To effectively implement (4), there are two obstacles to overcome:

- How to evaluate a multi-dimensional mutual information; and
- How to search the maximum.

Aiming at these problems, we have developed a gradient ascent optimization strategy to maximize MI in [9], [11], [12]. First, we show that a multi-dimensional MI can be decomposed into a series of one-dimensional MIs:

Let $\mathbf{x} = (X_1, X_2, \ldots, X_M)$ be a random vector representing the selected features $X_i, i = 1, 2, \ldots, M$, and Y the random variable corresponding class label. The mutual information between them can be written as:

$$I(\mathbf{x}, Y) = I\left(\left(X_1, X_2, \ldots, X_M\right), Y\right). \tag{5}$$

If $\mathbf{x}$ only has two components, i.e., $\mathbf{x} = (X_1, X_2)$, (5) becomes:

$$\begin{aligned} I(\mathbf{x}, Y) &= I((X_1, X_2), Y) \\ &= H(X_1, X_2) - H(X_1, X_2|Y). \end{aligned} \tag{6}$$

From (3), we can derive the following two equations:

$$H(X_1, X_2) = H(X_1) + H(X_2) - I(X_1, X_2), \tag{7}$$

and

$$H(X_1, X_2|Y) = H(X_1|Y) + H(X_2|Y) - I(X_1, X_2|Y). \tag{8}$$

Substituting $H(X_1, X_2)$ and $H(X_1, X_2|Y)$ of (6) into (7) and (8), we get:

$$\begin{aligned} I\left(\mathbf{x}, Y\right) &= H\left(X_1, X_2\right) - H\left(X_1, X_2|Y\right) \\ &= H\left(X_1\right) + H\left(X_2\right) - I\left(X_1, X_2\right) - H\left(X_1|Y\right) \\ &\quad - H\left(X_2|Y\right) + I\left(X_1, X_2|Y\right) \\ &= \sum_{i=1,2} I\left(X_i, Y\right) - I\left(X_1, X_2\right) + I\left(X_1, X_2|Y\right). \end{aligned} \tag{9}$$

Extending (9) to more than two components, we have the following equation:

$$\begin{aligned} I\left(\mathbf{x}, Y\right) &= \sum_i I\left(X_i, Y\right) - \sum_i \sum_{j>i} I\left(X_i, X_j\right) \\ &\quad + \sum_i \sum_{j>i} I\left(X_i, X_j|Y\right). \end{aligned} \tag{10}$$

According to (10), the mutual information between a vector $\mathbf{x}$ and a scaler Y can be decomposed by evaluating the MI between the component X_i and Y, and the MI between pairs of components X_i and X_j. All of them are one-dimensional mutual information, which can be effectively implemented.

To calculate one-dimensional MI, we can treat the amplitude of each frequency bin as a random variable X with continual value, and its category label as Y with discrete class labels (e.g., ω_1, ω_2, …, etc.) respectively. Thus, the MI between X and Y can be evaluated as follows (with a similar formula for $I(X_i, X_j)$):

$$\begin{aligned} I(X, Y) &= -\int_X p(\mathbf{x}) \log p(\mathbf{x})\, \mathrm{d}\mathbf{x} - \sum_y P(y) \log P(y) \\ &\quad + \sum_y \int_X p(\mathbf{x}, y) \log p(\mathbf{x}, y)\, \mathrm{d}\mathbf{x}. \end{aligned}$$

Meanwhile, based on equation (10), to maximize $I\left(\mathbf{x}, Y\right)$, $\mathbf{x} = (X_1, X_2, \ldots, X_M)$, the first variable can be chosen as:

$$X_1^0 = \max_i I\left(X_i, Y\right),$$

where X_k^0 represents the result of maximization at step k.

Then, the second variable is chosen as:

$$X_2^0 = \max_i \left[I(X_i, Y) - \sum_{X_i \neq X_1^0} I\left(X_i, X_1^0\right) + \sum_{X_i \neq X_1^0} I\left(X_i, X_1^0 | Y\right) \right].$$

The remaining variables are chosen in the same way until the pre-specified number, N, of variables is reached:

$$X_n^0 = \max_i \left[I(X_i, Y) - \sum_j \sum_{X_i \neq X_j^0} I\left(X_i, X_j^0\right) + \sum_j \sum_{X_i \neq X_j^0} I\left(X_i, X_j^0 | Y\right) \right].$$

where $X_j^0, j = 1, 2, \cdots, n-1$ are the variables already selected.

The above strategy selects features sequentially, and so avoids the problem of 'combinatorial explosion'. At each step, the next feature will be selected so as to maximize $I(\mathbf{x}, Y)$ incrementally. This is a similar idea to the gradient ascent or other hill-climbing algorithms.

Although we have shown that the key frequency features selected by mutual information can effectively provide useful discriminatory information, it is not recommended to completely replace the existing features, i.e., the harmonics features. This is because the new features are extracted purely on the discriminatory analysis. The amount of information extracted can be guaranteed, but the stability of the features is unsure. For example, the velocity change of vehicles is likely to affect the selected results. So the key frequency features should be better considered as a supplemental constituent to the major features, and a fusion approach should be applied to utilize both of them. As long as this strategy is followed, the final performance could be improved if the the key frequency components captured the new information, but will not degrade significantly even if they failed.

IV. Fusing acoustic feature sets

A natural way to combine the multiple feature sets for classification is by information fusion [10], [16], [17]. Two possible fusion strategies that can be applied for this task are feature level fusion and decision level fusion, which are discussed as follows.

A. Feature level fusion

The feature-level fusion is a medium-level fusion strategy, where some features extracted from raw data are combined for decision. Given the harmonics feature vector represented by

$$\mathbf{x}_h = \left\{ x_h^1, x_h^2, \cdots, x_h^M \right\},$$

where the superscribes represent different harmonic orders, and the key frequency feature vector denoted as

$$\mathbf{x}_k = \left\{ x_k^1, x_k^2, \cdots, x_k^N \right\},$$

where the superscribes represent different frequency bins, the feature level fusion can be simply implemented by concatenating the two sets of features, and the fused feature vector is formed as follows:

$$\mathbf{x}_{hk} = \left\{ x_h^1, x_h^2, \cdots, x_h^M, x_k^1, x_k^2, \cdots, x_k^N \right\}. \tag{11}$$

One of the aims of this research is to testify if the fusion of two set of features can improve classification accuracy. A fair assessment should be based on the feature vectors with the same dimensionality. Hence, the above fusion can be revised as:

$$\mathbf{x}'_{hk} = \left\{ x_h^1, x_h^2, \cdots, x_h^L, x_k^1, x_k^2, \cdots, x_k^K \right\}, \tag{12}$$

where $K + L = M$, and M is the dimensionality of the pre-specified harmonics feature space. The fused feature vectors now have the same dimensionality as the harmonic features', but with the L higher order (or other less important) harmonics replaced by the same number of key frequency components.

In this feature level fusion, since features from different extraction methods are augmented directly, a proper normalization should be applied to address the difference in the measurement scale.

According to our previous discussion, the fused feature vector $\mathbf{x}_{hk}$ or $\mathbf{x}'_{hk}$ tends to depict the acoustic signature more fully: the harmonics characterize the major noise sources and outline the global spectrum; the key frequency components provide other localized details of the spectrum.

The implementation of this feature level fusion is straightforward. However, one major problem associated with this fusion scheme is that a same classifier has to be applies to the fused feature set, which means that the two feature sets will be classified by the same classification algorithm. This is a unwanted consequence for this application, because according to Section II the two feature sets have different utilities and may have their individually favored classifiers. It is known that classification performance depends greatly on the characteristics of the data, and there is no single classifier that works best on all given data sets. Hence to achieve a better performance, the following decision level fusion is also investigated.

B. Decision level fusion

The decision level fusion is a high-level fusion, where separate intermediate decisions can be drawn from each individual features-set firstly and then combined to reach a global decision.

In pattern classification, choosing a suitable classifier for a given feature set is usually carried out by empirical tests. In this application, followed by the previous research [3], [5], we choose the multivariate Gaussian classifier (MGC) for the harmonic features. Currently-popular support vector machines (SVMs) [18], [19] have shown competitive performance with the best available algorithms in many classification areas, so were chosen as the classifiers for the key frequency component features. To combine the outputs from the classifiers MGC and SVM to reach a global decision, a Bayesian-based decision fusion is investigated, described as follows.

1) A modified Bayesian decision fusion method:

In the traditional Bayesian framework, several approaches are adopted to combine probabilistic information. Let $\mathbf{x}_i$, $i = 1, 2, \ldots, N$ be N information sources, and y the decision result, according to the *maximum a posteriori* (MAP) criterion, two usually-used fusion methods are listed as follows [10]:

$$p\left(y|\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\right) \propto \prod_{i=1}^{N} p\left(y|\mathbf{x}_i\right), \tag{13}$$

and

$$p\left(y|\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\right) \propto p\left(y\right) \prod_{i=1}^{N} p\left(\mathbf{x}_i|y\right). \tag{14}$$

It is known that both of the methods are based on certain independence assumptions. However in our application, the two feature sets are extracted from the frequency response of the same acoustic signal, and the independent assumption is very unlikely to hold. So directly applying the above fusion rules will result in larger discrepancy from the expected MAP result. To obtain a more accurate fusion, we propose the following improved fusion criterion.

First, based on our application scenario, two information sources, i.e., $\mathbf{x}_1$ and $\mathbf{x}_2$, are considered. According to the Bayes rule, the posterior probability can be written as follows:

$$p\left(y|\mathbf{x}_1, \mathbf{x}_2\right) = \frac{p(\mathbf{x}_1) p\left(y|\mathbf{x}_1\right) p\left(\mathbf{x}_2|y, \mathbf{x}_1\right)}{p\left(\mathbf{x}_1\right) p\left(\mathbf{x}_2|\mathbf{x}_1\right)}. \tag{15}$$

For the decision purpose, equation (15) can be simplified as:

$$\underset{y}{\operatorname{argmax}}\, p\left(y|\mathbf{x}_1, \mathbf{x}_2\right) \propto p\left(y|\mathbf{x}_1\right) p\left(\mathbf{x}_2|y, \mathbf{x}_1\right). \tag{16}$$

It can be found that (16) will be reduced to (14) if $\mathbf{x}_1$ is independent to $\mathbf{x}_2$.

To implement the fusion in (16), the conditional probabilities $p\left(y|\mathbf{x}_1\right)$ and $p\left(\mathbf{x}_2|y, \mathbf{x}_1\right)$ are needed. Through our previous discussion, the posterior $p\left(y|\mathbf{x}_1\right)$ can be effectively obtained from the SVM's output (see Section IV-B3 for details), and the likelihood $p\left(\mathbf{x}_2|y\right)$ can be conveniently obtained from the outputs of the MGC. Then a major problem is to estimate $p\left(\mathbf{x}_2|y, \mathbf{x}_1\right)$ based on all available information, which can be formulated as follows:

$$p_{\mathbf{x}_k}\left(\mathbf{x}_h|y\right) \doteq \hat{p}\left(\mathbf{x}_2|y, \mathbf{x}_1\right) \leftarrow \left\{p\left(\mathbf{x}_2|y\right), \mathbf{x}_1, \mathbf{x}_2\right\} \tag{17}$$

So, according to our specific application a more accurate MAP decision rule can be re-written as:

$$\begin{aligned} \underset{y}{\operatorname{argmax}}\, p\left(y|\mathbf{x}_h, \mathbf{x}_k\right) &\propto p\left(y|\mathbf{x}_k\right) p\left(\mathbf{x}_h|y, \mathbf{x}_k\right) \\ &\approx p\left(y|\mathbf{x}_k\right) p_{\mathbf{x}_k}\left(\mathbf{x}_h|y\right), \end{aligned} \tag{18}$$

where $\mathbf{x}_h$ and $\mathbf{x}_k$ represent the harmonics features and the key frequency features respectively; $p_{\mathbf{x}_k}\left(\mathbf{x}_h|y\right) = \hat{p}\left(\mathbf{x}_h|y, \mathbf{x}_k\right)$ is an estimate of $p\left(\mathbf{x}_h|y, \mathbf{x}_k\right)$ given knowledge of $p\left(\mathbf{x}_h|y\right)$ and $\mathbf{x}_k$. To get $p_{\mathbf{x}_k}\left(\mathbf{x}_h|y\right)$, we propose an approach based on a simple information-theoretical criterion, presented as follow.

2) Modulating multi-dimensional Gaussian distribution:

The previous research [3], [5] has shown that the multi-dimensional Gaussian distribution is an effective estimation for the probability density of the harmonic features. So let $\mathbf{x}_h$ be a d-dimensional harmonics feature vector, the likelihood function will be:

$$p(\mathbf{x}_h|y) = \frac{1}{(2\pi)^{d/2} \left|\Sigma\right|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}_h - \mu)^{\top} \Sigma^{-1} (\mathbf{x}_h - \mu)\right), \tag{19}$$

where μ and Σ are the mean vector and the covariance matrix respectively.

Because the feature vector $\mathbf{x}_h$ is not independent to $\mathbf{x}_k$, the appearance of $\mathbf{x}_k$ will reduce the uncertainty of $\mathbf{x}_h$. In information theory, the uncertainty is usually measured by entropy. So to estimate $p\left(\mathbf{x}_h|y, \mathbf{x}_k\right)$ from $p\left(\mathbf{x}_h|y\right)$, one convenient approach (based on the above basic information theorem) is to reduce the entropy of $p\left(\mathbf{x}_h|y\right)$ by including the knowledge of $\mathbf{x}_k$. It is known that the entropy of (19) is:

$$\ln\left(\sqrt{(2\,\pi\, e)^{d} \left|\Sigma\right|}\right). \tag{20}$$

To reduce (20), we may modulate the covariance matrix Σ by:

$$\Sigma^{*} = (1 - \beta)\, \Sigma, \tag{21}$$

where Σ^{*} is the updated covariance matrix, and β is a modulation factor, decided by $\mathbf{x}_k$. Using the Gaussian function, β can be defined as follows:

$$\beta = \exp\left\{-\gamma \|\mathbf{x}_h - \mathbf{x}_k\|\right\}, \tag{22}$$

where coefficient γ controls the depth of modulation, and its suitable value can be empirically decided by cross-validations.

Thus, the conditional probability $p\left(\mathbf{x}_h|y, \mathbf{x}_k\right)$ is estimated by modulating the covariance matrix Σ of $p\left(\mathbf{x}_h|y\right)$, i.e.,

$$\hat{p}\left(\mathbf{x}_h|y, \mathbf{x}_k\right) \sim \mathcal{N}(\mu, \Sigma^{*}), \tag{23}$$

where Σ^{*} is the updated covariance matrix, including the correlation between $\mathbf{x}_k$ and $\mathbf{x}_h$ (see (22)).

3) Calibrating SVMs' output to probability:

After obtaining the likelihood from MGC, to implement the fusion approach in (18) we still need another posterior probability from the SVM classifier, i.e., $p\left(y|\mathbf{x}_k\right)$. However, the Standard SVMs do not provide posterior probability. To find this probability, one of the convenient ways is by training an additional sigmoid function to approximate a posterior probability [20]. To explain the method, we need to firstly introduce several necessary SVM formulas [21], [22].

Let $\mathbf{x}_i$ be a feature (data) vector, $y_i \in (+1, -1)$ be the class label, $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_L)$, be the Lagrange multipliers, L

the number of examples and b a threshold. The SVM classifier can be represented as:

$$f(\mathbf{x}) = \sum_{i=1}^{L} y_i \alpha_i K(\mathbf{x_i}, \mathbf{x}) + b,$$

where $K(\mathbf{x}, \mathbf{x}') = \Phi(\mathbf{x})^T \Phi(\mathbf{x}')$ is an appropriate kernel function which has a corresponding inner product expansion, Φ. The commonly-used functions are polynomials and Gaussian radial basis functions (RBFs):

$$K(\mathbf{x}, \mathbf{x}') = \left(\mathbf{x}^T \mathbf{x}' + 1\right)^d, \tag{24}$$

and

$$K(\mathbf{x}, \mathbf{x}') = \exp\left\{\frac{-\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma}\right\}. \tag{25}$$

To get the posterior probability, we applied a mapping method introduced in [20], where an additional sigmoid function is used to approximate the necessary posterior probability. In detail, the posterior probability is trained by a sigmoid function:

$$p(y|\mathbf{x}) \approx \frac{1}{1 + \exp(Af(\mathbf{x}) + B)}, \tag{26}$$

where parameters A and B are found by minimizing the following cross-entropy error function:

$$\operatorname*{argmin}_{A,B} \left[-\sum_{i=1}^{L} t_i \log(p(y|\mathbf{x}_i)) + (1 - t_i) \log(1 - p(y|\mathbf{x}_i))\right], \tag{27}$$

with $t_i = \frac{y_i + 1}{2}$. The details on the calculation of (27) can be found in [20].

Based on the above discussions, a decision fusion approach can be implemented, summarized as follows:

- A SVM is used to draw a decision based on the key frequency feature vector $\mathbf{x}_k$;
- A maximum likelihood classifier, i.e. MGC, is applied to the harmonic features vector $\mathbf{x}_h$; and
- The improved fusion rule, proposed in (11), is then used to achieve the final global decision.

Comparing with other Bayesian fusion rules, e.g., (13) and (14), the proposed method does not need the independence assumption, and is based on a more accurate MAP criterion (see (11)). Meanwhile, benefiting from the specific characters of the application data (e.g, the Multivariate Gaussian distribution for harmonic features), its implementation is simplified, avoiding those more complicated methods, such as Bayesian inferences.

V. Experimental results

To assess the proposed information fusion approach, experiments are carried out based on a multi-category vehicles acoustic data set from US ARL [3]. The ARL data set consists of recoded acoustic signals from five types of ground vehicles, named as $\mathbf{V1_t}$, $\mathbf{V2_t}$, $\mathbf{V3_w}$, $\mathbf{V4_w}$, and $\mathbf{V5_w}$ (the subscript 't' or 'w' stands for the tracked vehicles or wheeled vehicles respectively). These vehicles run 6 cycles around a prearranged track at different time, and the corresponding acoustic signals are recorded for the assessment.

Table I
The number of runs and the total sample numbers for five types of vehicles: tracked vehicles $\mathbf{V1_t}$ and $\mathbf{V2_t}$; wheeled vehicles $\mathbf{V3_w}$, $\mathbf{V4_w}$ and $\mathbf{V5_w}$.

Vehicle Class	Number of Runs	Total Number of Samples
$\mathbf{V1_t}$	6	1734
$\mathbf{V2_t}$	6	4230
$\mathbf{V3_w}$	6	5154
$\mathbf{V4_w}$	6	2358
$\mathbf{V5_w}$	6	2698

To obtain frequency domain representation, Fourier transform (FFT) is firstly applied to each second of acoustic signal with Hamming window, and the output of the spectral data (i.e., a 351 dimensional frequency domain vector $\mathbf{x}$) is considered as one of the samples for these five vehicles. Then feature extraction is carried out on the sample $\mathbf{x}$ to get the two sets of features, i.e., the harmonics feature vector $\mathbf{x}_h$ and the key frequency feature vector $\mathbf{x}_k$. Subsequently, these feature vectors are fed into the classifier(s), and the final classification result will be obtained from the fusion algorithms.

The type label and the total number of spectral vectors for each vehicle are summarized in Table I. A 'run' is assumed to correspond to a vehicle moving a $360°$ circle around the track, and a sample means the FFT result at one second time interval. Differences in the total numbers of samples reflect the vehicles' different moving speeds.

As we discussed in Section IV, the features to be fused are came from the harmonic extraction and mutual information evaluation respectively. The left column of Figure V illustrates the 351 dimensional spectral vectors for the five types of vehicles (corresponding to $\mathbf{V1_t}$ - $\mathbf{V5_w}$, from top to bottom). For each type of vehicle, 20 samples are illustrated in Figure V, reflecting the variations at different sampling time and different runs. The right column of Figure V shows the 21 dimensional harmonic features extracted from the above spectral vectors for these five vehicles. The amplitudes of these harmonics will form a harmonic feature vector $\mathbf{x}_h = \left\{x_h^1, x_h^2, \cdots, x_h^{21}\right\}$.

From Figure V, it can be seen than the spectral responses of vehicle's sounds are quiet complex, consisting of many formants that did not appear at the exact positions of the integral multipliers of the fundamental frequency. There are also severe within-class variations in the acoustic features (see the extracted harmonics features). As for the between-class variations, there are many overlapped formants among 5 vehicles. For examples, Figure V(a) and (g) have similar peaks around frequency 50 Hz and 100 Hz; Figure V(g) and (i) show similar frequency response between 1-50 Hz. These evidences show that vehicle noises are much more complex, and a single feature set may not be able to cover all of the acoustic characters.

In the experiments for accuracy comparison, half of runs for

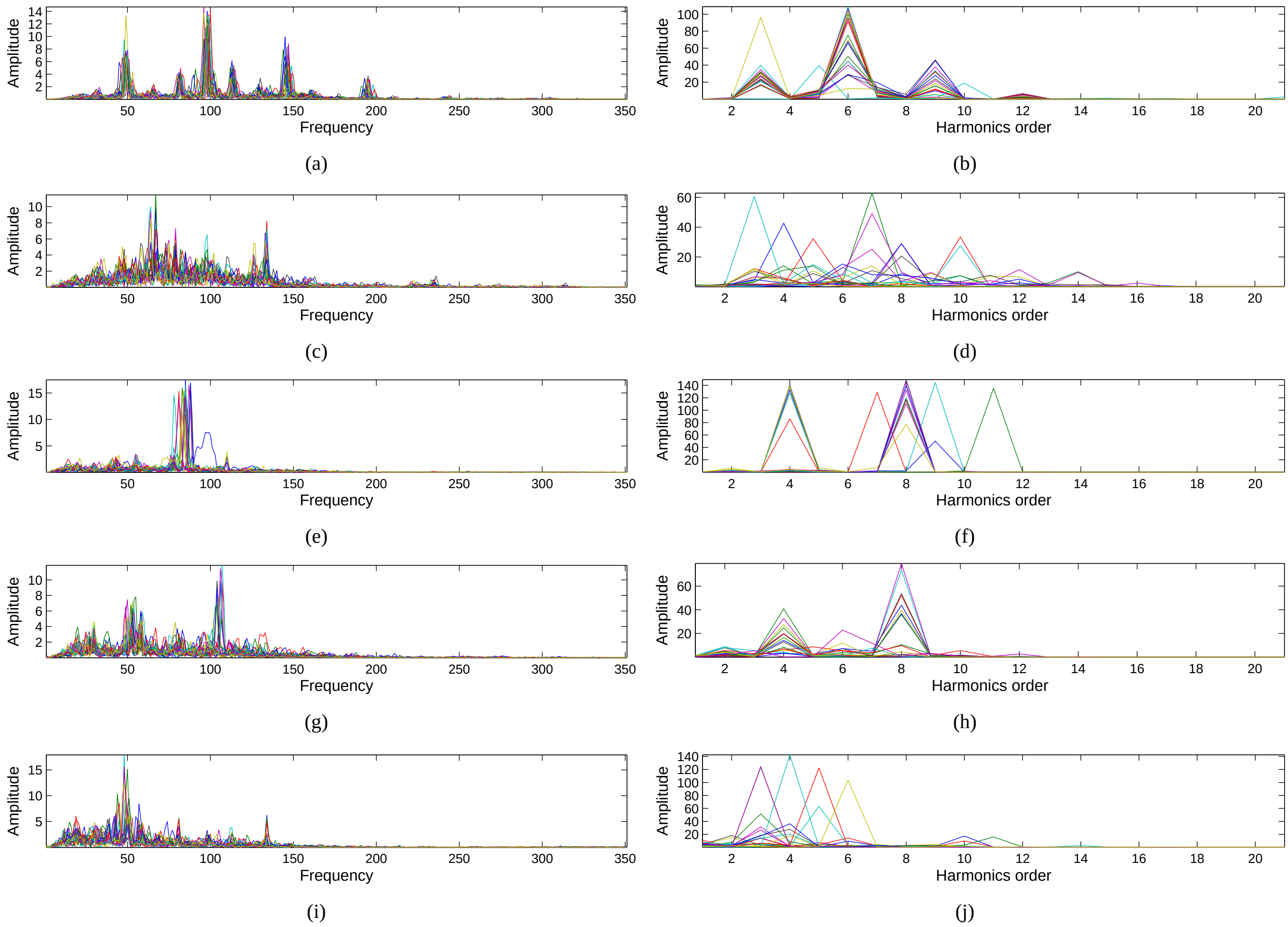

Figure 2. Illustration of spectrum (left column) and harmonic features (right column) for five vehicles $\mathbf{V1_t}$ (top) - $\mathbf{V5_w}$ (bottom), respectively; 20 samples (depicted in different colors) for each class.

each vehicle (i.e., 3 runs) were randomly chosen to estimate the statistical parameters for feature extraction, such as the harmonic features' means vector μ, covariance matrix Σ and mutual information I. The remaining half runs form the test set on which performance was assessed. Next, feature extraction are carried out based on the methods introduced in Section IV. Following the results in [3], the harmonic number is chosen as 21[1].

As we discussed previously, SVMs [18], [19], [21], [22] and Multivariate Gaussian classifier (MGC) [5] were chosen as the classifiers in these experiments. Because SVMs are inherently binary (two-class) classifiers, $\binom{5}{2}$ one-against-one classifiers were used with subsequent majority voting to give a multi-class result. The kernel function used is an inhomogeneous polynomial. The penalty parameter C is tested between 10^{-3} and 10^{5}, and polynomial order is tested from 1–10 by a two fold validation procedure using only training data. The polynomial order 3 and $C = 20$ were finally found as the best values for this SVM, and applied to the following testing stage. The training data are also used to estimate the mean vector and covariance matrix for MGC.

To avoid bias on random samplings, the testing was repeated 10 times to allow an estimate of the error inherent in this sampling process. The 10 times classification results based on different feature sets are then shown in Figure 3.

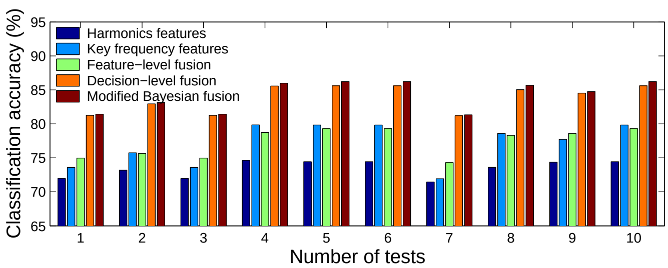

Figure 3. Comparison of classification accuracy for different feature sets and fusion methods; 10 times tests with random chosen 3 runs for training and the remaining 3 runs for testing; the accuracy is the overall result for all 5 type of vehicles.

[1]The main reason to choose the harmonics number of 21 is to keep consistent with the previous studies [3]. However, we note that there may be a minimum sufficient number for harmonics but that will depend on different applications.

Table II
MEAN CLASSIFICATION ACCURACY OF 10 TIME TESTS.

Methods	Average accuracy (%)
Harmonics feature set	73.44
Key frequency feature set	77.05
Feature-level fusion	77.34
Decision-level fusion	83.86
Modified Bayesian decision fusion	84.24

From Figure 3, the following results are observed:

- For each individual feature sets, the key frequency feature set (the second column) achieved better classification accuracy than the the harmonics feature set (the first column). This testified the effectiveness of using mutual information for feature extraction introduced in Section III;
- The feature level fusion (the third column) is slightly better than using each individual feature sets (the first and the second column) but is very close to the best result from each individual sources. This phenomenon has been observed in previous sensor fusion research literature;
- The decision level fusion (the fourth and fifth column) achieved significant improvements than those using each individual feature sets (the first and second column), and are also much better than the feature level fusion (the third column); This demonstrated the efficacy of the proposed information approach; and
- The improvement of the modified fusion method (proposed in Section IV-B) is found consistently in all of 10 times tests (see the fifth column).

The average numbers for the above 10 times tests' results are summarized in Table II, which further demonstrated the effectiveness of proposed fusion methods.

VI. CONCLUSIONS

In this paper, we developed an information fusion approach for acoustic ground vehicle classification. First, we argued that multiple feature sets are needed to improve the vehicles classification accuracy. Then, a key frequency feature vector is added to the harmonic feature vector, to amend the ignored discriminatory information. Finally, a modified Bayesian decision fusion was proposed to better combine the two sets of features. Experiments were carried out to assess the classification accuracies of the fusion approach, based on a multi-category vehicles acoustic data set. The results showed that significant improvement of classification accuracy has been achieved by the fusion approach. Future research will address the features' stability with regard to vehicles' velocity changes, and extended this approach to other more complicated data sets.

ACKNOWLEDGMENTS

We gratefully acknowledge support by ARL/MoD through IBM on the Information Technology Alliance research programme.

REFERENCES

[1] H. Wu, M. Siegel, and P. Khosla. Vehicle sound signature recognition by frequency vector principal component analysis. *IEEE Transactions on Instrument and Measurement*, 48(5):1005–1009, 1999.
[2] M. Duarte and Y. H. Hu. Vehicle classification in distributed sensor networks. *Journal of Parallel and Distributed Computing*, 64:826–838, 2004.
[3] T. R. Damarla and G. Whipps. Multiple target tracking and classification improvement using data fusion at node level using acoustic signals. Technical Report, ARL.
[4] H. Wu and J.M. Mendel. Classification of battlefield ground vehicles using acoustic features and fuzzy logic rule-based classifiers. *IEEE Transactions on Fuzzy Systems*, 15(1):56–72, 2007.
[5] T. R. Damarla, T. Pham, and D. Lake. An algorithm for classifying multiple targets using acoustic signature. In *Proceedings of SPIE Signal Processing, Sensor Fusion and Target Recognition*, 2004.
[6] D. Lake. Harmonic phase coupling for battlefield acoustic target identification. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1998.
[7] D. Lake. Tracking fundamental frequency for synchronous mechanical diagnostic signal processing. In *Proceedings of 9th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, 1998.
[8] R. Battiti. Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on Neural Networks*, 5(4):537–550, July 1994.
[9] Baofeng Guo, S. Gunn, R. I. Damper, and J. Nelson. Adaptive band selection for hyperspectral image classification using mutual information. In *Proceedings of Eighth International Conference of Information Fusion*, pages 630–637, Philadelphia, PA, USA, 2005.
[10] J. Manyika and H. Durrant-Whyte. *Data Fusion and Sensor Management: A Decentralized Information-Theoretic Approach*. Ellis Horwood, New-York, London, 1994.
[11] Baofeng Guo, S.R. Gunn, R.I. Damper, and J.D.B. Nelson. Band selection for hyperspectral image classification using mutual information. *IEEE Geoscience and Remote Sensing Letters*, 4(3):522–526, 2007.
[12] Baofeng Guo and M. Nixon. Gait feature subset selection by mutual information. In *Proceedings of the First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, Washington, U.S., 2007.
[13] K. Torkkola. Feature extraction by non-parametric mutual information maximization. *Journal of Machine Learning Research*, 3:1415–1438, 2003.
[14] N. Kwak and C-H. Choi. Input feature selection by mutual information based on parzen window. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24(12):1667–1671, 2002.
[15] P. Viola and W. Wells. Alignment by maximization of mutual information. In *Proceedings of 5th International Conference on Computer Vision*, pages 16–23, 1995.
[16] G. V. Veres, M. S. Nixon, L. Middleton, and J. N. Carter. Fusion of dynamic and static features for gait recognition over time. In *Proceedings of Eighth International Conference of Information Fusion*, pages –, Philadelphia, PA, USA, 2005.
[17] A. I. Bazin and M. S. Nixon. Probabilistic combination of static and dynamic gait features for verification. In *Proceedings of Biometric Technology for Human Identification II, SPIE Defense and Security Symposium 5779*, Orlando (Kissimmee), Florida USA, 2005.
[18] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152, Pittsburgh, Pennsylvania, United States, 1992.
[19] C. Cortes and V. N. Vapnik. Support-vector networks. *Machine Learning*, 20(3):1–25, 1995.
[20] J. Platt. *Advances in large margin classifiers*, chapter Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. Cambridge: MIT Press, 2000.
[21] C. Burges. A tutorial on support vector machines for pattern recognition. *Knowledge Discovery and Data Mining*, 2(2):121–167, 1998.
[22] V. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999, September 1999.