12th International Conference on Information Fusion
Seattle, WA, USA, July 6-9, 2009

Acoustic Vehicle Classification by Fusing with Semantic Annotation

Baofeng Guo
HANGZHOU DIANZI UNIV.
Hangzhou, 310018, CHINA
bguouk@gmail.com

Mark. S. Nixon
Electronics and Computer Sci.
University of Southampton
Southampton SO17 1BJ, UK
msn@ecs.soton.ac.uk

Thyagaraju Damarla
Army Research Laboratory
2800 Powdermill Road
Adelphi, MD, USA
rdamarla@arl.army.mil

Abstract – *Current research on acoustic vehicle classification has been generally aimed at utilizing various feature extraction methods and pattern recognition techniques. Previous research in gait biometrics has shown that domain knowledge or semantic enrichment can assist in improving the classification accuracy. In this paper, we address the problem of semantic enrichment by learning the semantic attributes from the training set, and then formalize the domain knowledge by using ontologies. We first consider a simple data ontology, and discuss how to use it for classification. Next we propose a scheme, which uses a semantic attribute to mediate information fusion for acoustic vehicle classification. To assess the proposed approaches, experiments are carried out based on a data set containing acoustic signals from five types of vehicles. Results indicate that whether the above semantic enrichment can lead to improvement depends on the accuracy of semantic annotation. Among the two enrichment schemes, semantically mediated information fusion achieves less significant improvement, but is insensitive to the annotation error.*

Keywords: Acoustic vehicle classification, semantic enrichment, information fusion.

1 Introduction

Acoustic sensors, such as a microphone array, can collect an aeroacoustic signal (i.e., passive acoustic signal) to identify the type and localize the position of a working ground vehicle. Acoustic sensors can be used in sensor networks for applications such as traffic monitoring and battlefield surveillance [1, 2]. They become more and more attractive because they can be rapidly deployed and have low cost [3, 4]. One of the research areas in acoustic sensor processing is to identify the type of vehicle [1, 5], which can help improve the performance of tracking [6].

Previous approaches on acoustic vehicle classification mainly focus on signal processing and pattern recognition techniques. Many acoustic features can be extracted to classify working ground vehicles. The commonly-used features include moment measurements [1], eigenvectors [7], linear prediction coefficients [8], Mel frequency cepstral coefficients [9], the levels of various harmonics [5, 10]. Among them, harmonic features have achieved good classification performance [5, 10, 11], with a stable and compact feature representation.

Apart from the above typical features that extract a vehicle's information at the signal level, domain experts may use human descriptions of what has been heard. These *semantic* words can connect with some high level descriptions regarding the studied vehicles, such as the engine volume, the tracked or wheeled vehicle, the size of the vehicle, etc, which suggests the possibility of exploiting this domain knowledge or semantic representation for classification. Moreover due to cheap cost and low energy consumption, acoustic sensors can be easily deployed in multiple places to collect signals of interest from different locations. Therefore, the requirements of integration and communication among different sensor nodes become overwhelming. In this case, semantic enrichment is an appealing approach to alleviate the glut of too much data which lacks compact information.

Previous research in gait biometrics [12] used human labeled semantic attributes regarding descriptions of human appearance for classification purposes. In this research, we seek to automatically extract the vehicles' semantic attributes to enhance decision making processes, rather than simply augmenting the semantic information with existing acoustic features. The framework of our approach is shown in Fig. 1. Here, we classify the vehicle that generated the sound recorded by microphones. In a conventional pattern recognition framework (the left side of the dotted line), features are extracted from the sensor data and these features are filtered according to perceived information content, prior to use in classification. We can enrich this process by semantic data (the right side of the dotted line), which we shall represent using ontologies. These processes, embedded with the vehicles' domain knowledge elicited by experts in the form of ontologies, will contribute to the data fusion processes which can lead to the combined and enriched decision.

In web technologies, the use of ontologies is usual, but in acoustic vehicle classification a problem arises because

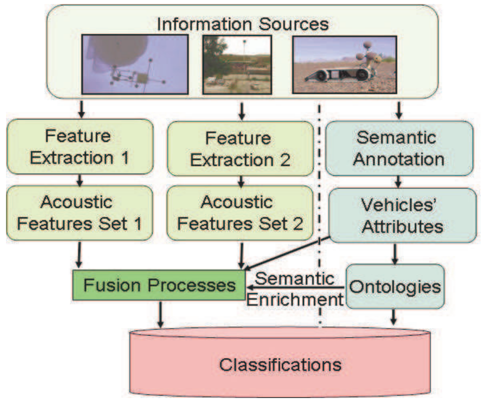


Figure 1: Illustration of semantic enrichment scheme for acoustic vehicle classification

the received signal is not at the same level as the semantic interpretation. It is therefore necessary to find the correspondence between low-level features which can be automatically extracted from the acoustic signal and the semantic concepts used in the ontology. In this research, we refer to this task as *semantic annotation* (more specifically, relate to the received acoustic signals to certain semantic descriptions, such as size, engine volume, wheel information etc.) and implement it in a supervised manner learning the semantic annotations from the data, thus automatically labeling the data. Then we consider using an ontology for acoustic vehicle classification, as well as applying semantic attributes to mediate information fusion.

The rest of this paper is organized as follows. In Section 2, we discuss semantic representation and reasoning within the acoustic data. Next in Section 3, we discuss how to use the semantic attribute to mediate the fusion proportion for acoustic vehicle classification. Experimental results are presented in Section 4. Finally, we end this paper with conclusions and future proposals.

2 Vehicle classification by semantic enrichment

The semantic enrichment for the acoustic vehicle classification can be carried out by identifying semantic concepts and their relations appearing in the studied acoustic data set. This procedure has a large coverage of different ontologies, such as sensor ontology (semantic description of the sensors;), sequence ontology (semantic description of events detected), data ontology (semantic description of data received), and supporting ontology (semantic description of concepts that would effect all three mentioned ontologies.) [13]. All these ontologies can have a potential to make contributions to the vehicle classification. For example, sensor ontologies can help focus more on reliable data, such as when the event "Condition A met, and sensor B is likely to receive corrupted signals" is detected. Based on the current acoustic data available, the data ontology is the most feasible option to implement, because it totally depends on the features extracted and how wide the semantic description will cover. So in this research we focus mainly on the vehicle related classes and properties, such as "wheel information", "weight" and "size", whereas other relevant properties, such as environmental factors will be studied in the future.

In detail, we initially consider the simplest data ontology, which involves only one semantic description, i.e. a vehicle's wheel information (i.e., the transport mechanism, whether it is tire, track, runner, etc). This attribute is relatively easy to be detected from the signal received by the sensor. This toy example might appear to be naive, but it makes the whole demonstration complete and allows us to determine if any improvement on classification accuracy can be achieved after adopting semantic enrichment.

Although study in semantics has made explicit claims concerning the representation of each meaning regarding the studied domain by different words, the relation between signal and semantic attributes and its structures are often left implicit. So when we are considering how to represent acoustical data semantically, two fundamental questions may arise, namely:

1. How is the semantic representation related to the actual signal?

2. How are the meanings of different concepts related to one another?

Since we have training data from each type of vehicle that can be labeled by a specific semantic concept, we can model the first problem as supervised learning. For example, for the semantic concept "a vehicle's wheel information", we can separate the training data into two groups: the air tire vehicles and the tracked vehicles. Then a binary classifier can be trained to detect this concept, and be applied to the test signals, which will be annotated to the presence or absence of the concept.

For the second question, we can consider an ontology, which defines a set of concepts, their characteristics and their relations to each other. These definitions allow us to describe and to use reasoning on the studied domain. A naive vehicle ontology for this particular acoustic data set is illustrated as in Fig. 2. This simple ontology has a three level structure, and uses one semantic attribute (i.e., the vehicle's wheel information). The acoustic data can then be enriched by this semantic meaning as it includes certain vehicle domain knowledge. In this way, the acoustic features are likely to be better separated thereby improving classification capability.

This simple scheme uses semantic attributes and ontology in a straightforward manner. However, there is a risk regarding this methodology to improve the classification accuracy. Based on our previous discussion, the classification in Fig. 2 actually involves three classifiers, where the first binary classifier annotates the semantic attribute (the tracked or tire label) to each data sample, and the second and the third classifier further separate each individual vehicle from the

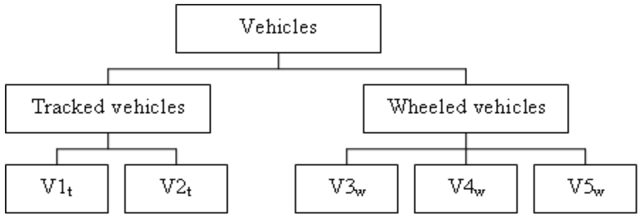


Figure 2: Illustration of a simple ontology for the acoustic vehicle data set

tire and tracked vehicle group. It can be found that the use of this ontology can improve classification accuracy. This can be interpreted by an intuitive understanding of "divide and conquer", or more specifically by an assumption that the classifier separating less numbers of classes will give more accurate result than those separating more numbers of classes. Apart from the fact that there is no rigorous proof of this claim, it is apparent that the annotation error in the first level will pass on to both of the second and the third classifier, which may, on the other hand, deteriorates rather than improves the classification accuracy. Therefore, in this research we are not only using ontology directly but also exploiting the semantic attributes in another way, i.e., to mediate the process of data fusion, which is presented in the next section.

3 Acoustic information fusion mediated by semantic attributions

In previous research such as in [12], the relevant semantic attributes have been labeled manually to augment the existing features. In order to improve this scheme, we need to exploit the semantic meaning regarding the acoustic data *automatically*, and then enable reasoning about it in a framework that can be aligned with data fusion. In this section, we discuss using multiple feature sets for acoustic vehicle classification, and give a simple example showing how the semantic attributes can be used to mediate a probabilistic based fusion.

3.1 Multiple feature sets for acoustic vehicle classification

The acoustic signal of a working vehicle is complicated. It is well known that the vehicle's sound may come from multiple sources, not exclusively from the engine, but also from exhaust, tires, gears, etc [14–16]. Classification based on one extracted feature set is therefore likely to be confined by its assumed sound production model, and can only efficiently capture one of the many aspects of the acoustic signature. Although it could be argued that this model can target the major attributes and makes the extracted features represent the most important acoustic knowledge, given the intricate nature of the vehicles' sounds it is still likely to lose information, especially when the assumed model is not comprehensive. For example, in a harmonic oscillator model it is difficult to represent the non-harmonic elements, which can also contribute significantly to the desired acoustic signature [15].

To handle the above problem, multiple feature sets may be used to classify the vehicle. For example, in our previous research [15], we address this problem from the perspectives of joint "generative-discriminative" feature extraction and information fusion. In detail, we first categorize the multiple vehicle noises into two groups based on their resonant properties, which leads to the subsequent "generative-discriminative" feature extraction and a probabilistic fusion framework.

The applied feature extraction methods, where global and detailed spectrum information can be obtained together, produce two feature sets respectively. The first set of features we used is the amplitudes of a series of harmonics components. This feature-set, characterizing the acoustic factors related to the fundamental frequency of resonance, has a clear physical origin and can be represented effectively by a "generative" Gaussian model. The second set of features are named as *key frequency components*, designated to reflect other minor (in the sense of sound loudness or energy in some circumstances) but also important (in the sense of discriminatory capability) acoustic characters, such as tires' friction noise, aerodynamic noise, etc. Because of the compound origins of these features (e.g., involved with the multiple sound production sources), they are better extracted by a discriminative analysis to avoid modeling each source of sound production separately. To search for the key frequency components, mutual information (MI), a metric based on the statistical dependence between two random variables, is applied. Selection of the key acoustic features by the mutual information can help to retain those frequency components (in this research, we mainly consider the frequency domain representation of a vehicle's acoustic signal) that contribute most to the discriminatory information, meeting our goal of fusing information for classification.

In associated with this feature extraction, information fusion is introduced to combine the acoustic knowledge represented by the above two sets of features, as well as their different underlying sound production. In this sense, information fusion can be achieved not only by combining different *sources* of data, such as in the traditional sensor fusion, but also by different feature extraction or "experts", which can compensate for the deficiency in *model assumptions* or *knowledge acquisitions*. A typical Bayesian fusion rule (for two feature sets) can be represented as:

$$p\left(y|\mathbf{x}_1,\mathbf{x}_2\right) \propto p\left(y\right)\prod_{i=1}^{2} p\left(\mathbf{x}_i|y\right). \qquad (1)$$

Assuming the same prior probability and applying $\log$ to (1), we get a sum fusion rule as follows:

$$\log p\left(y|\mathbf{x}_1,\mathbf{x}_2\right) \propto \log p\left(\mathbf{x}_1|y\right) + \log p\left(\mathbf{x}_2|y\right). \qquad (2)$$

3.2 Fusion driven by semantic attributions

In information fusion, an ideal combination rule should be adaptive to the factors that can affect the final fusion performance. In acoustic data fusion, the following factors should be taken into account:

1. Feature set's capability to capture the desired acoustic signature

   Information fusion involves multiple data sources or feature sets naturally. In our application, many acoustic factors can represent various aspects of an acoustic signature, and each feature set, either based on different sensors or different model assumptions, can be used to characterize these factors. Meanwhile, each of these feature sets has different capability or functionality to represent the desired acoustic signature, as well as has different contribution to the classification accuracy of working ground vehicles. For example in our case, the first set of features aims to represent internal sound production (e.g., the engine noise), and the second set of features is extracted to account for the sound production from the vehicle's exterior parts (i.e., the tire friction noise and air turbulence noise). Here, the engine noise is the dominant constituent of the overall vehicular loudness during the majority of time. On the contrary, the tire friction noise and air turbulence noise are more volatile. For example, changes of velocity will severely affect air turbulence noise, and change of the road condition is very likely to influence the tire friction noise. Therefore for this specific application, the amount of information extracted by the second feature-set is unstable. If this difference of feature set's capability is taken into account in the designing of the specific fusion rule, a better performance could be expected, e.g., by increasing the weights of the reliable feature sets and reducing the contribution of the weaker feature sets.

2. The quality of the feature extraction

   Before the fusion of information, each feature set has to be extracted from the data by a specific algorithm. The feature extraction algorithm usually involves some parameters estimation and parameters choice problems, which will affect the quality of the features extracted.

   For example, in this research the first set of features is a group of harmonic components, extracted by the fundamental frequency and the peak detection algorithms. Here, how to choose the optimal number of the harmonics to correctly characterize the engine's formants is not straightforward. This is because the variability of engine types and their resonance characteristics. To include more harmonics may introduce redundancy and cause some problems for the following classification algorithm (e.g., to calculate the inverse of the covariance matrix in the multivariate Gaussian classifier). On the other hand, a smaller number of harmonics may risk the classification accuracy due to insufficient representation of the engine noise.

   The second set of features is extracted based on a computationally effective discriminatory analysis, and a group of key frequency components is selected by Mutual Information (MI). The MI based feature extraction also needs to estimate the statistical properties of the training data, and the accuracy of this learning will directly affect the capability of the selected features.

   These two examples show that the quality of feature extraction may be different based on different sets of parameters. Therefore, how to reflect the quality of the extracted feature sets is another problem, which should be considered in the fusion rule.

3. Application scenarios and other factors

   In acoustic vehicle classification, the received acoustic signal will be affected by many ambient factors such as temperature, wind speed, humidity, etc, as well as some operating conditions such as vehicle distance to the sensors, vehicle load, surround buildings, etc. All these factors can change the accuracy of the assumed sound production models, and then the quality of the extracted feature sets. So considering these factors into the fusion procedure may also lead to improvement of performance. For example the air turbulence noise will become quite trivial if the vehicle is far away from the sensors, but the harmonic feature could keep almost the same effectiveness in this case. Therefore, if the distance information can be correctly used in the fusion processing, e.g, to put more emphasis on the harmonic features in the above scenario, it is likely to improve the classification performance.

We have discussed some factors that can affect the information fusion performance. Now we argue that the semantic attributes, i.e., high level domain knowledge, can help to describe some of the factors. To describe a vehicle, we can use different levels of concepts. For example, at the signal level, we can use the frequency representation of the received acoustic signal to characterize a vehicle; at the information level, we can induce the statistics of the features for this vehicle; and at the knowledge level, we may describe this vehicle using some human understandable concepts such as size, carriage, weight, etc. Conventional techniques are mainly focusing on the signal and the information level descriptions, but to information fusion, the knowledge level description, i.e., the semantic attributes, can provide valuable clues to improve performance.

As we discussed before, fusion performance can be improved if the fusion rule can correctly address the capability of each source of information, e.g., to give the more powerful feature set a bigger weight in the fusion formulation. In this research, suppose we know the vehicle's wheel information (i.e, the tire or track), we can then use this semantic attribute to improve the fusion rule. Intuitively, if the vehicle has tires, we can conjecture that its friction

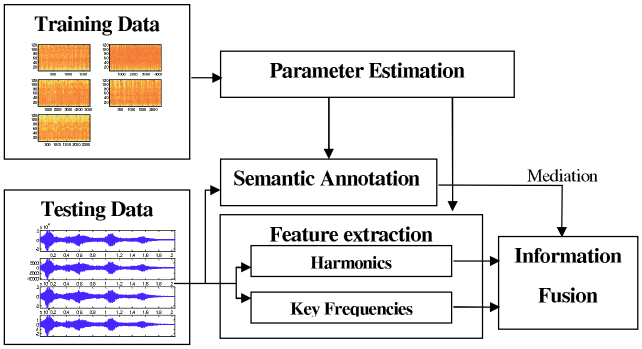


Figure 3: Semantically mediated acoustic information fusion

noise with the road would be much less influential than a tracked vehicle. Therefore, in the fusion procedure, we should reduce the contribution of the feature set representing the tire friction noise. Moreover, if we know that the size of this vehicle is big and its weight is heavy, then we may figure out the engine type of this vehicle roughly. This may tell us how accurate to use the harmonic oscillator to model this type of engine, and then give the fusion rule an indication what kind of confidence should be assigned to the harmonic features.

A typical weighted sum decision rule can be described as follows [17]:

$$C_{sum}\left(\mathbf{x}_1, \mathbf{x}_2, \alpha\right) = \alpha C_1\left(\mathbf{x}_1\right) + (1-\alpha) C_2\left(\mathbf{x}_2\right) \tag{3}$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are two feature sets, α the fusion weight (or fusion proportion), and $C_{(\cdot)}$ the classification functions. If we can link the semantic attributes with the fusion weight α, the high level domain knowledge will be embedded in this fusion procedure implicitly.

Based on the above discussion, the high level domain knowledge, e.g., semantic attributes, can be found useful to mediate the data fusion. A diagram of exploiting semantic attributes in this research is illustrated in Fig. 3, where a new module of semantic annotation is added and its result, i.e., a detected semantic attribute, will be used to adjust the fusion weight in the fusion rule, e.g., α in (3).

To implement the scheme described in Fig. 3, we first need to automatically extract semantic descriptors from acoustic signals. This semantic annotation can be posed as a problem of either supervised or unsupervised learning. In the case of the supervised learning, we can collect a set of training signals with and without the concept of interest, and train a binary classifier to detect this concept. The classifier was then applied to the unseen testing signals, which were annotated with respect to the presence or absence of this concept.

To demonstrate how to implement this semantically mediated data fusion, we give a simple example based on one semantic attribute related with this research. Given a binary classifier

$$C(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \text{ is a tracked vehicle} \\ -1 & \text{if } \mathbf{x} \text{ is a vehicle withairtires} \end{cases} \tag{4}$$

which is trained by the training data to detect the wheel information of the vehicle. Let $L(\mathbf{x})$ be the number of the components of the feature vector $\mathbf{x}$, we can use the semantic attribute detected by $C(\mathbf{x})$ to control the fusion proportion of each information source, such as:

$$L(\mathbf{x}_1) = \begin{cases} m & \text{if } C(\mathbf{x}) = 1 \\ n & \text{if } C(\mathbf{x}) = -1 \end{cases} \tag{5}$$

and

$$L(\mathbf{x}_2) = \begin{cases} N-m & \text{if } C(\mathbf{x}) = 1 \\ N-n & \text{if } C(\mathbf{x}) = -1 \end{cases} \tag{6}$$

where N stands for the total number of features, which is constrained by some application factors, such as computational load of the sensor network, communication bandwidth, etc. This scheme, i.e., adjusting the components number of each feature set according to the detected semantic attribute, can be found consistent with the traditional fusion rule, such as (3).

Given $\mathbf{x} = (x^1, x^2, \ldots, x^k)$ and $\mathbf{x}' = (x^1, x^2, \ldots, x^k, x^{(k+1)}, \ldots, x^{(k+l)})$, we have

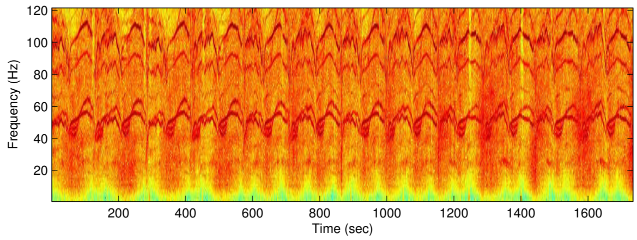
$$\begin{aligned} p(\mathbf{x}) &= \int_{x^{(k+1)}} \cdots \int_{x^{(k+l)}} p(\mathbf{x}') \mathrm{d}x^{(k+1)} \cdots \mathrm{d}x^{(k+l)} \\ &\geq p(\mathbf{x}') \end{aligned} \tag{7}$$

Therefore, (7) shows that changing the dimensionality of the feature vector will lead to a different probability and then finally change the fusion proportion in the fusion rules, such as in (2). This is also similar to the traditional weighted fusion rule in (3). In (5) and (6), we indicate that dimensionality of each feature set may change according to their different semantic label. However, the detailed relation between the semantic label and the dimensionality, i.e., the value of m and n, is left implicit. Currently there are no methods available to deduce these numbers theoretically, so we consider using the training data to learn these parameters empirically.

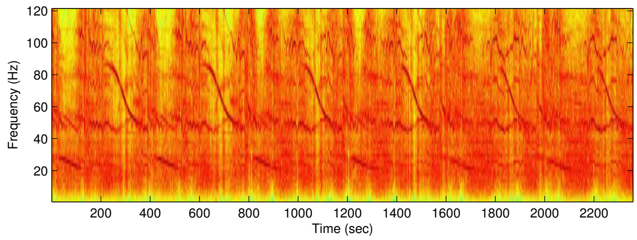
4 Simulation results

To assess the proposed approaches, simulations are carried out based on a multi-category vehicles acoustic data set from US ARL [6]. 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’ denotes the tracked or wheeled vehicles, respectively). These vehicles cover 6 running-cycles around a prearranged track separately, and the corresponding acoustic signals are recorded by a microphone array for the assessment (see examples of acoustic signals in Fig. 4).

To obtain a frequency domain representation, the Fourier transform (FFT) is first applied to each second of the



(a) Time-Frequency response for a tracked vehicle

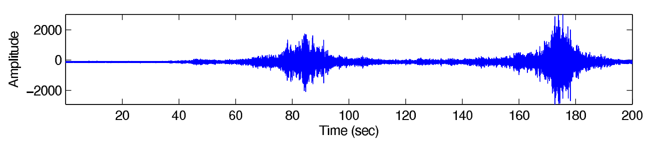


(b) Time-Frequency response for a vehicle with tire

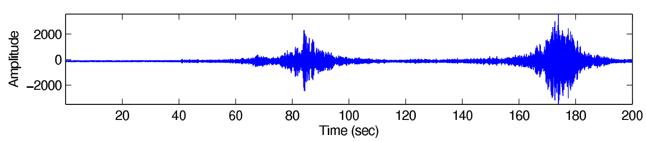
Figure 5: Examples of training data

Table 2: Annotation accuracy (%) for the carriage attribute based on two classifiers

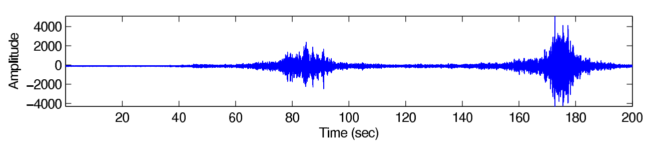
Classifiers	Strong classifier	Weaker classifier
Annotation accuracy	96.5	87.4



(a) Microphone #1



(b) Microphone #3



(c) Microphone #5

Figure 4: Examples of acoustic signals (20 sec) from a microphone array

Table 1: 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

acoustic data with a Hamming window, and the output of the spectral data (a 351 dimensional frequency domain vector $\mathbf{x}$) is considered as one of the samples for these five vehicles. Fig. 5 shows time-frequency response for two different kinds of vehicles, which are used as training data. The type label and the total number of the (spectral) data vectors for each vehicle are summarized in Table 1. A "run" corresponds to a vehicle moving a 360° circle around the track and the sensors array, and a sample means the FFT result at one second signal.

In the simulations, half of the runs from each vehicle (i.e., 3 runs from all 6 runs) were randomly chosen as the training data, and the remaining half forms the test set. In this data set, each run consists of about 290 to 860 seconds of acoustic data depending on vehicles' different running speeds. Tests are carried out based on each second of the acoustic data (i.e., to classify the vehicles in each second interval, which is useful for vehicle tracking) from the 3 test runs, and the overall accuracies are summarized from the above results for all 5 types of vehicles.

As discussed previously, for semantically enrichment the first step is to apply classifiers to automatically extract the semantic attributes from the acoustic data. In this research to observe the influence of annotation accuracy, we test two annotation classifiers: the first one is a SVM (Support Vector Machine) with the polynomial kernel order 2 and penalty coefficient $C = 0.1$ (these parameters were chosen by a validation using the training data), and the second one is a multivariate Gaussian classifier (MGC) [6]. Because SVMs are less affected by the dimensionality of input, 121 dimensional FFT acoustic data is directly input to this classifier to achieve higher accuracy. On the other hand, the multivariate Gaussian classifier uses the lower 21 dimensional harmonic features as the input. Here the SVM is considered as a "stronger" classifier (i.e., expected higher classification accuracy) than the MGC (a "weaker classifier") in that it uses much higher dimensionality input and more complex learning. Based on the training set, we use the above two classifiers to annotate the vehicle's wheel attribute (i.e, the air tire or the track information), and then calculate the annotation accuracy, listed in Table 2. It is seen that the stronger SVM classifier indeed achieves much higher annotation accuracy (96.5%) than the weaker MGC classifier (87.4%), at the cost of higher dimension input and prolonged training and parameters selection.

After using these two classifiers to annotate the acoustic data, we test the classification accuracy through the semantic enrichment described in Fig. 2. Based on this scheme, three individual classifiers are applied: the first is the annotation classifier to detect the vehicle wheel information, the second classifier separates the tracked vehicles into two detailed categories: $\mathbf{V1_t}$ and $\mathbf{V2_t}$, and the third classifier separates the vehicles with air tires into three types: $\mathbf{V3_w}$, $\mathbf{V4_w}$ and $\mathbf{V5_w}$. The features fed into the second and the third

Table 3: Classification accuracy (%) based on semantic enrichment for the carriage attribute

Methods	Direct classification	Ontology-based weaker classifier	Ontology-based stronger classifier
Accuracy	73.4	71.9	79.1

Table 4: Classification accuracy (%) based on semantically mediated fusion

Methods	Direct fusion	Semantically mediated fusion	Ontology based fusion
Stronger classifier	83.3	84.18	85.3
Weaker classifier	83.3	84.23	77.4

classifiers are the 21 dimensional harmonic features.

Test results are listed in Table 3, where we can see that using the stronger classifier (with 96.5% annotation accuracy), the classification accuracy is improved significantly from 73.4% to 79.1%, but using the weaker classifier (with 87.4% annotation accuracy), the classification accuracy is deteriorated from 73.4% to 71.9%. This simple scheme exploits the data structure from the point view of an ontology, but it is actually equivalent to the standard "divide and conquer" strategy. If this separation, in other words the semantic annotation in this case, is accurate, less class confusion will occur in the next stage classification because a small number of classes is involved at the third level of Fig. 2. This leads to the improvement of the classification accuracy, as evidenced in Table 3. But if the first step of separation is not accurate, all the annotation misclassification will pass on to the classifiers in the next stage. When these errors offset all the benefits brought by the less class confusion, the classification accuracy will be degraded.

The motivation of applying semantic enrichment is to use domain knowledge, and one of the key issues is how to exploit this domain knowledge. Possible approaches are either relying on other information sources, such as new sensors or human intelligence, or based on exploring the existing training data, such as the above example of semantic annotation. The ontology used in this example is actually like a clustering pre-processing in the name of the semantic annotation. The results in Table 3 show that whether improvement can be achieved depends on how accurate the semantic attributes can be obtained.

To test the semantically mediated data fusion, the second feature set, i.e., the key frequency component feature, are extracted based on the method introduced in [15]. Both the dimensionality of the harmonic feature set and the key frequency component feature set are chosen as 21, i.e., they have initially the same contributions in the fusion processing. In the test, the parameters that decide the fusion proportions of the two feature sets, i.e., the numbers m and n in (5) and (6), are estimated as $m = 6$ and $n = 4$ using the training data, respectively.

Three methods are compared in Table 4. The first method directly uses the fusion method introduced in [15], the second method uses the simple semantically mediated fusion described in (5) and (6), and the third method uses the ontology described in Fig 2. The difference between the second and the third method is that the former one uses the semantic attribute to control the fusion proportions of the two feature sets but the latter one uses the semantic attribute to separate the data in the first place. From Table 4, it is seen that when the semantic annotation is accurate (using the stronger classifier), the best classification result is achieved by the ontology based method. Meanwhile, if the semantic annotation is not very accurate (using the weaker classifier), the worst classification result is also achieved by the ontology based method. On the other hand, the semantically mediated fusion achieves slight improvement in both of the cases. One possible explanation is that in this case the annotation misclassification error is passed on to the fusion proportions rather than next stage classifiers that are more sensitive to such errors.

Although the improvement of the semantically mediated fusion is found consistently in every random tests, the increase of classification accuracy is trivial. This result can also be predicted by observing the trained two parameters m and n. There is no big difference between $m = 6$ and $n = 4$, which means less necessity to adjust fusion proportion based on this semantic attribute (i.e. carriage information). However the result of this initial test should not be extended to other semantic attributes, which are still very likely to give useful indication on how to control the fusion proportions, such as the use of semantic human description in gait biometric [12]. Future research to test other semantic attributes will be carried out when more suitable data set and descriptions are available.

5 Conclusions

In this paper, we have discussed how to automatically extract the vehicle semantic attribute and formalized it as an ontology for vehicle classification. We considered two implementation schemes: 1) ontology based classification that explicitly use semantic attribute, and 2) semantically mediated data fusion that implicitly use semantic attributes. The simulation results have shown that the ontology-based vehicle classification can improve classification accuracy given the semantic attributes were accurately annotated. If the semantic attributes are not accurately annotated, the error will propagate to the next level of classification, and then finally leads to deteriorate results. The Semantic-mediated fusion achieved slight improvement with weak dependence on the accuracy of semantic annotation. However, these conclusions are based on a single semantic attribute and a naive data-driven ontology. This research represents an initial study in this new area. We have shown that semantic annotations can be learned from the data, and enrich the classification and the fusion process. Future research will consider more vehicle attributes and more realistic ontologies. In this research, we use "semantic" to describe the meaning attached with the high level domain knowledge (vehicle category) for classification. However, one may argue that only humans or minds are the appropriate entities, which can define these attributes. Therefore, this research was exploring an alternative way to "automatically" extract the semantic attributes by machines, which may not exactly match the definition used in other research areas.

Acknowledgments

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

The authors would like to thank the reviewers for their constructive comments.

References

[1] D.W. Thomas and B.R. Wilkins. The analysis of vehicle sounds for recognition. *Pattern Recognition*, 4(4):379–389, 1972.

[2] Z. Sun, G. Bebis, and R. Milleri. On-road vehicle detection: A review. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 28(5):694–711, 2006.

[3] M. Duarte and Y. H. Hu. Vehicle classification in distributed sensor networks. *Journal of Parallel and Distributed Computing*, 64:826–838, 2004.

[4] J. Hsieh, S. Yu, Y. Chen, and W. Hu. Automatic traffic surveillance system for vehicle tracking and classification. *IEEE Transaction on Intelligent Transportation Systems*, 7(2):175–187, 2006.

[5] T. 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*, pages 421–427, 2004.

[6] T. Damarla and G. Whipps. Multiple target tracking and classification improvement using data fusion at node level using acoustic signals. In *Proceedings of SPIE Unattended ground sensor technologies and applications*, volume 5796, pages 19–27, 2005.

[7] 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.

[8] A. Y. Nooralahiyan, H. R. Kirby, and D. McKeown. Vehicle classification by acoustic signature. *Mathematical and Computer Modelling*, 27(9-11):205–214, 1998.

[9] M. Munich. Bayesian subspace methods for acoustic signature recognition. In *Proceedings of the 12th European Signal Processing Conference*, pages –, Vienna, Austria, 2004.

[10] D. Lake. Harmonic phase coupling for battlefield acoustic target identification. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 2049–2052, 1998.

[11] 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.

[12] S. Samangooei, M. S. Nixon, and B. Guo. The use of semantic human description as a soft biometric. In *Proceeding of Biometrics: Theory, Applications, and Systems 2008*, pages 1–7, 2008.

[13] B. Guo, Y. Wang, P. Smart, N. Shadbolt, M.S Nixon, and R. Damarla. Approaching semantically-mediated acoustic data fusion. In *Proceeding of Military Communications Conference (MILCOM)*, pages 1–7, 2007.

[14] V. Cevher, R. Chellappa, and J. McClellan. Joint acoustic-video fingerprinting of vehicles, part i. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing 2007*, volume 2, pages 745–748, 2007.

[15] B. Guo, M. S. Nixon, and R. Damarla. Acoustic information fusion for ground vehicle classification. In *Proceeding of The 11th International Conference of Information Fusion*, pages 1–7, 2008.

[16] V. Cevher, R. Chellappa, and J. McClellan. Vehicle speed estimation using acoustic wave patterns. *IEEE Transactions on Signal Processing*, 57(1):30–47, Jan. 2009.

[17] R. Hu and R.I. Damper. A 'no panacea theorem' for classifier combination. *Pattern Recognition*, 41(8):2665–2673, 2008.