



Acoustic Target Classification In Wireless Audio–Sensor Networks

Baljeet Malhotra, Ioanis Nikolaidis, and Janelle Harms

Department of Computing Science,
University of Alberta,
Edmonton, Alberta, Canada T6G 2E8
Emails: {baljeet,yannis,harms}@cs.ualberta.ca

Abstract. Target tracking is one of the important problems in wireless sensor networks. We consider an aspect of tracking: the classification of targets based on the acoustic signals produced by vehicles. We present a naïve classifier and simple distributed schemes based on features extracted from acoustic signals. We demonstrate a novel way of using Aura matrices to create a new feature derived from the Power Spectral Density (PSD) of a signal, which performs at par with other existing features. An experimental study has been conducted using real acoustic signals of different vehicles in an urban setting. Our proposed schemes using a naïve classifier achieved highly accurate results in classifying different vehicles into two classes. Communication overheads were also computed to capture the tradeoff of energy cost vs. classification quality.

1 Introduction

Networked sensors can be equipped with various sensing devices, as well as memory, processor, radio, and a power supply. However, they are still constrained by limited memory, processing power, channel capacity, and, most importantly, energy reserves. When tracking is considered as an application, data-intensive sources (e.g., high frame rate / high resolution video) is usually avoided as being more energy expensive than low data rate sources. For this reason tracking using audio signals is usually preferable. Vehicle tracking on acoustic data is based on the fact that different vehicles produce distinctly different acoustic signals because their engine and propulsion mechanisms are unique [9]. The problem of vehicle detection using the acoustic signature has existed for years and many solutions have been proposed in the literature [1, 9, 10]. Recently, target classification based on acoustic signals in wireless sensor networks has been addressed in [3, 6]. The advantage of sensor networks is that they provide redundancy in terms of sensing and processing units. Hence, they can operate together in a distributed coordinated fashion to detect and report the presence of a target vehicle, possibly refining the tracking and classification quality as the target is moving.

We should add that vehicle tracking includes various objectives that must be supported by a number of steps. These steps include vehicle detection, identification and/or classification, and localization. Depending on the specific tracking

objectives, all or combination of these steps may be required. In this paper we restrict our attention to classification. Classification is necessary because sensors can report on a specific moving vehicle only after they recognize vehicles that are of some interest. Classification is naturally more challenging if there are multiple targets of various types (e.g., tanks, jeeps, other types of military vehicles, civilian vehicles, etc.). Furthermore, there may be a number of vehicles of the same type *e.g.* tanks of a particular make. We define as *classification* the problem of identifying which class a vehicle belongs to. Identifying a *particular* vehicle goes one step further and is not within the scope of the current paper.

Various techniques have been proposed to address the classification problem [1, 3, 9], relying on *feature extraction* that differ in the way features are extracted. For example, [1], proposed a wavelet based method for feature extraction, which works as following: three different types of acoustic signature are extracted: (i) squeak sound, (ii) sound under motion, and (iii) exhaust sound. The data points contained by each of these signatures are decomposed to a 12 element feature vector using the multi-resolution analysis [7]. These 12 element feature vectors represent the energy concentration of the signature signal at 12 different resolution levels. The continuous wavelet transform (CWT) and the STFT (short-time Fourier transform) plots were used for two other feature vectors. Finally, these feature vectors are used to compute the distance between the reference and unknown signatures. Wu *et. al.* in [10] proposed a principle component analysis based method for recognition of acoustic signatures. The basic idea of their proposed method is to use together the mean adjusted sound spectrum, and key eigenvalues of the covariance matrix to characterize an acoustic signature.

An adaptive threshold based algorithm is proposed in [2] for vehicle detection, based on the average energy of an acoustic signal crossing a threshold value before a decision on the detection of the vehicle can be made. The threshold is updated adaptively. Also, [3] details experiments carried out during the 3rd SensIT situational experiment (SITEX02) organized by DARPA/IXO. Various military vehicles were used in these experiments, real world data was generated and archived for future studies. The objective was to detect and accurately locate vehicles using energy-based localization algorithm. Frequency spectrum based features are extracted from the acoustic and seismic signals captured by the sensors. These features were extracted by using 512-point FFT. Then, for classification purposes, three different approaches are used *i.e.* (i) k-nearest-neighbor classifier (ii) Maximum Likelihood classifier, and (iii) support-vector-machine classifier. In [6], a framework for collaborative signal processing in sensor networks is designed for the purpose of multiple targets detection, classification, and tracking.

In this paper we study the impact of increasing the complexity and memory footprint of a classification algorithm to achieve better classification accuracy. We consider this to be a reasonable tradeoff since what is usually assumed to be expensive in terms of energy is communication, and not computation/storage (within reason of course). While we are also increasing the computation cost, we argue that as long as the computation is allowed to be completed within a

reasonable amount of time, computation can be spread over a longer period of time by proper reduction of the CPU clock [13].

One distinct contributions of our work is that in order to maximally benefit from the information collected by a sensor, we consider multiple representations of features. Some, are well known (FFT, PSD, etc.) but we also introduce *Aura matrices*. Aura matrices [4] have been used in the past for analyzing and predicting texture patterns [11]. In our study we create “artificial” 2-dimensional “textures” by arranging PSD data into matrices, and then using Aura matrices to summarize the information of the 2-dimensional matrix. In other words, Aura matrices attempt to visually approximate the arrangement of PSD values. For details about the construction of Aura matrices the reader is referred to [4].

Section 2 presents the features considered in this study. Section 3 describes a naïve classification scheme and elementary forms for a distributed implementation over a sensor network. Section 4 presents performance evaluation results in terms of classification accuracy and energy expenditure tradeoffs for the different distributed implementations. Finally, Section 5 summarizes the findings of the paper and outlines future research objectives.

2 Features Extraction

A vehicle sound is a stochastic signal. In practice, sound of a moving vehicle observed over a period of time will not be a stationary signal. However, a signal of fairly short duration can be treated as a stationary signal [10]. In our case we chose the signal’s duration to be 23.27 ms *i.e.* 512 data points sampled at a frequency of 22 kHz. In our study we considered five features that are generated using FFT and PSD of time series data of a given signal. In the rest of the paper the terms signal and signature are used interchangeably.

1. **Linear FFT feature (LFFT):** This feature is generated using FFT of 512 data points that gave us a linear vector (of size 512) representing frequencies with a resolution of 21.48 Hz.
2. **Linear PSD feature (LPSD):** This feature is generated by taking Power Spectral Density (PSD) estimates of 512 data points using the Thomson multi-taper method. With a resolution of 21.48 Hz that method gave a linear vector (of size 512) to form a linear PSD feature.
3. **Multidimensional FFT feature (MFFT):** In this case 50 blocks of 512 FFT data points are used to form a multidimensional FFT feature. This feature can be seen as a matrix of size 512×50 . The size 50 was determined by trial and error method using our acoustic data set.
4. **Multidimensional PSD feature (MPSD):** In this case 50 blocks of 512 PSD data points are used to form a multidimensional PSD based feature. This feature can be seen as a matrix of size 512×50 .
5. **Aura of a Multidimensional PSD feature (AMPSD):** It has been shown in [5] that power spectral density (PSD) is not an optimal feature for signal recognition. We explored to improve the PSD based feature using

some established statistical techniques, namely Aura matrices. In order to construct AMPSD features we simply compute Aura of a MPSD matrix. For computing the Aura of a matrix the reader is referred to [4].

The first four features are popular in the signal classification literature. The fifth feature (AMPSD), which is part of our contributions has been added to improve the performance of MPSD feature.

3 A Naïve Classifier

Existing techniques such as k -Nearest Neighbor (k -NN) can be used by sensor nodes to perform the classification. It is based on the idea that similar objects are closer to each other in a multidimensional feature space. k -NN is one of the simplest, yet accurate, classification methods and recently it has been used in sensor networks for target classification [3, 6]. Unfortunately, finding k for the optimum solution is non-trivial. In contrast to k -NN algorithm we adopt a naïve approach where first a learning set $|U|$ is defined for each class of vehicles. Equal number of samples are assumed to be in the learning set of these classes. Note that class labels of the learning samples are known in advance. Now in order to classify an unknown sample using a particular feature, the naïve classifier does the following: For each class, and for each sample of the $|U|$ samples in the learning set of each class, we calculate the distance of the unknown sample from the learning set samples. We determine the average distance of the unknown sample from the learning set samples of each class. The unknown sample is determined to be in the class with the smallest average distance.

As a vehicle crosses through an area monitored by a sensor network, the nodes self-organize in neighborhood “clusters” using a technique similar to [12] but where tie breaking criteria for selection of the “master” node is the signal quality of the monitored vehicle (the master is the node with higher average power of received signal, hence possibly closest to the vehicle). Each of the neighborhoods can perform a classification method independently of other neighborhoods. However, multiple neighborhoods may collaborate with each other for two main reasons: (i) better accuracy in classifying a vehicle, and (ii) sharing the costs associated with classification. In our study we examined various scenarios of single and multiple neighborhoods based classification. We propose four basic schemes:

- **Single Neighborhood Based Classification using Local Signatures (SNBC-LS):** In this scheme each sensor in a neighborhood uses its own local copy of vehicle’s signature. Sensors compute the distance of unknown sample from known samples in each of the class using their own (local) copy of unknown sample. The master node collects results from all participating nodes in the neighborhood and classifies the unknown sample (vehicle).
- **Single Neighborhood Based Classification using Global Signatures (SNBC-GS):** In this scheme each sensor in a neighborhood uses a global copy of vehicle’s signature. A global copy of vehicle’s signature is transmitted

to them by the master node of a neighborhood. Sensors compute the distance of unknown sample from the known sample in each of the class using the global copy of the unknown sample. The master node collects results from all participating nodes in the neighborhood and classifies the unknown sample (vehicle).

- **Multiple Neighborhood Based Classification using Local Signatures (MNBC-LS):** In this scheme a master node not only collects results from sensors in its own neighborhood, but it also invokes its adjoining neighborhoods to seek the classification results. All sensors in participating neighborhoods use their local copy of vehicle’s signature.
- **Multiple Neighborhood Based Classification using Global Signatures (MNBC-GS):** The basic difference between this scheme and the previous scheme (MNBC-LS) is that sensors in a particular neighborhood use a global copy of vehicle’s signature provided to them by their respective master nodes.

4 Experimental Study

We used acoustic signals of various urban ground vehicles. These acoustics signals were recorded using Panasonic US395 microphone. Approximately 50 samples of various vehicles were recorded at two main locations of Bonnie-Doon mall and University of Alberta bus stop in the city of Edmonton. This real database of acoustic samples consisted of acoustic signatures of ETS buses, which are part of the public transportation system at the city of Edmonton, different types of cars, small trucks, SUVs, and mini vans. All acoustic signals were first recorded to the local memory of the microphone, and then transferred to MATLAB for the simulation of classification algorithms.

We assume that every sensor has a copy of learning set, U for each class. A sensor’s copy of unknown (testing) sample is different from others because of their geographical positioning. In order to create a local copy of unknown (testing) sample for a sensor we attenuate the signal based on its distance from the moving vehicle and add noise to it. A vehicle’s sound can also be degraded by reverberations, however, we considered an outdoor open environment, so we have neglected the effect of reverberations. In our simulation we considered various scenarios for sensor setup. In these experiments sensors are assumed to be placed along two straight parallel lines, i.e., as they would be deployed along the sides of a street. Sensors are placed 5m apart and their sensing and radio range is 15m. Vehicle is considered to be moving with a speed of $53 \pm$ Km/H.

4.1 Performance Metrics

We consider two performance metrics: (i) classification accuracy, and (ii) energy expenditure. Classification accuracy is computed based on the leave-one-out policy. Under this policy one sample is removed from the acoustic dataset consisting of all samples in a class. This sample is called the testing sample. The rest of the

samples in the dataset constitute the learning set U for that class. Class label of the testing sample is assumed to be unknown. Then, the distance of the testing sample is computed from all the samples in the learning set of each class. This process is repeated for all samples in our acoustic dataset. If two samples are represented by matrices $X_{m \times n}$ and $X'_{m \times n}$, then the distance between them is computed as following:

$$d = \sum_{q=1}^{|n|} \sum_{p=1}^{|m|} \sqrt{(x_{pq} - x'_{pq})^2} \quad (1)$$

Classification accuracy is calculated as a percentage of testing samples that are correctly classified from the total number of testing samples. In the experiments we used a simplistic case of only two classes. The objective was to classify the previously mentioned vehicles into two classes *i.e.* ETS buses and other vehicles that are not ETS buses.

4.2 SNBC-LS & SNBC-GS

In the SNBC-LS scheme all sensors use their own copy of vehicle's signature. Note that every sensor's copy might be different, because of its position. In these experiments we attenuated a sensor's signal based on its distance from the moving vehicle. In the case of SNBC-GS scheme, all sensors use a common copy of an unknown vehicle's signature that is provided to them by the master node.

The results for classification accuracy in SNBC-LS and SNBC-GS schemes are presented in Figures 1(a) and 1(b), respectively. With a better copy of vehicle's signal available to all participating sensors, SNBC-GS scheme performs better than SNBC-LS. Classification using LFFT, LPSD, MPSD and AMPSD based features improved in SNBC-GS scheme. As noticed in these figures classification accuracy improves as the number of sensors per neighborhood increases. The reason for improved accuracy is that collecting distance measurements from all sensors in a neighborhood, and taking their average improves the approximation of the distance measurements. An accuracy of 99% is achieved using MFFT feature that is better than the accuracies reported in [8], which uses k -NN classifier.

Communication costs are much higher in the SNBC-GS scheme. The results for SNBC-LS and SNBC-GS schemes are presented in the Figures 2(a) and 2(b), respectively. The reason for larger costs in SNBC-GS are obvious. The master node needs to communicate the signature in its neighborhood. More specifically it transmits 512 data point in the case of LFFT and LMSD features and 512×50 data points for MFFT, MPSD and AMPSD features.

These results suggest that there is a trade-off between communication costs, computational costs, and achieving a higher classification accuracy. A higher classification accuracy comes at a higher energy cost for sensors. Note that when number of training samples per sensor are fixed, then varying the number of sensors per neighborhood affects the classification accuracy. Having more sensors in a neighborhood increases the classification accuracy but at a higher cost

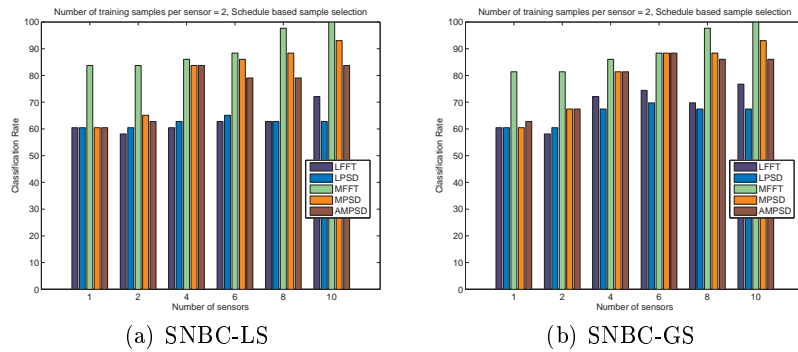


Fig. 1. Classification accuracy.

of communication. On the other hand selection of a feature is also an important decision. Some features are more expensive to use than others, but their classification results are better.

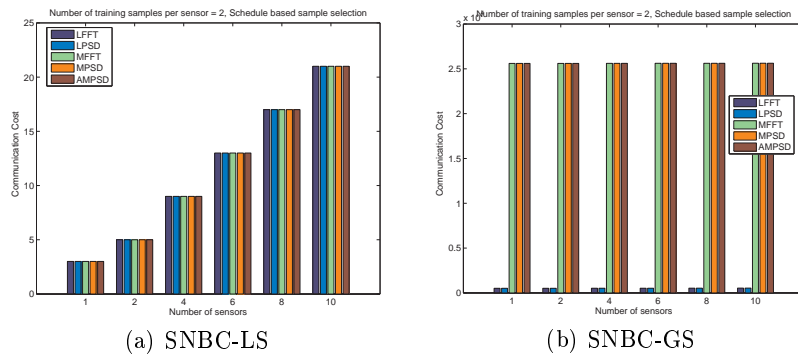


Fig. 2. Communication costs.

4.3 MNBC-LS & MNBC-GS

Classification based on multiple neighborhoods may arise in many situations. Consider the case in which number of nodes (and samples) in a particular neighborhood are not enough to achieve a higher classification accuracy. This may happen due to a number of reasons like malfunctioning of sensors because of insufficient battery power or physical damages. In these experiments we assume that only two sensors are available per neighborhood. We also assume that learning set U is distributed over multiple neighborhoods though the sensors in a single neighborhood have a common learning set. In the MNBC-LS scheme all

sensors in the corresponding neighborhood use their own copy of vehicle's signature. In the case of MNBC-GS scheme all sensors use a common copy provided to them by the master node of their neighborhood.

The results for classification accuracy in MNBC-LS and MNBC-GS schemes are presented in the Figures 3(a) and 3(b), respectively. There is not much difference between classification results for single neighborhood scheme (*e.g.* SNBC-LS) and multiple neighborhood scheme (*e.g.* MNBC-LS). The reason for this behavior is that in multiple neighborhood schemes, learning samples are distributed over multiple neighborhoods that result in a smaller learning set for each of the neighborhoods. Therefore distance calculation of unknown sample is approximated from readings of multiple neighborhoods instead of sensors from a single neighborhood. In the case of global signatures, a similar trend can be seen in multiple neighborhood schemes. With a better copy of vehicle's signal available to all participating sensors in a neighborhood, MNBC-GS scheme performs better than MNBC-LS. Again, classification accuracy using LFFT, LPSD, MPSD and AMPSD based features improved in MNBC-GS scheme.

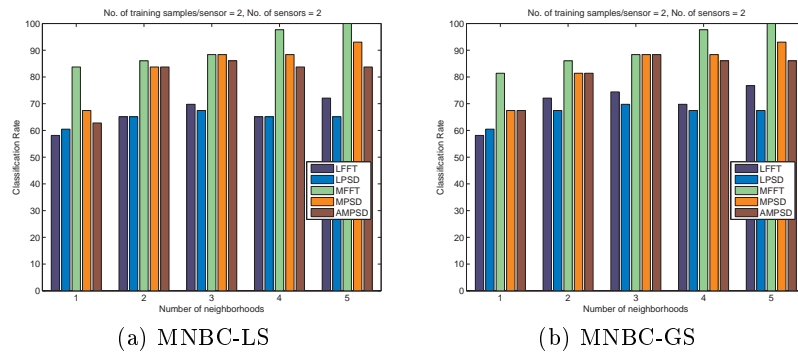


Fig. 3. Classification accuracy.

As expected communication costs are much higher in MNBC-GS scheme. These results are presented in the Figures 4(a) and 4(b) for MNBC-LS and MNBC-GS schemes respectively. As mentioned previously MNBC-GS is more expensive because the master node transmits 512 data point in the case of LFFT and LMSD features and 512×50 data points for MFFT, MPSD and AMPSD features.

5 Conclusions and Future Directions

Classifying ground vehicles is an important problem in wireless sensor networks. Features extracted from the acoustic signatures of these vehicles form the basis for classification. Whether sensor networks provide for efficient implementation of tracking, depends on whether necessary operations, such as classification, can

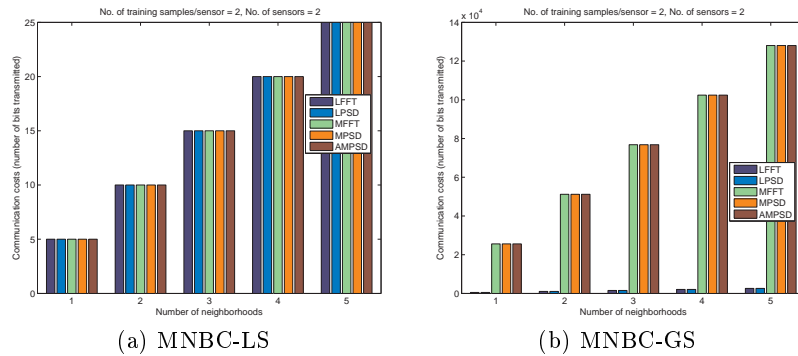


Fig. 4. Communication costs.

be performed efficiently in a distributed fashion, achieving high classification accuracy at a reasonable energy cost. In this paper we proposed several distributed schemes for vehicle classification. These schemes are based on the idea of collaborations in a single and multiple sensor neighborhoods. We conducted a simulation study using real acoustic signals of urban ground vehicles. Simulation results have revealed the performance of our proposed schemes. Our proposed schemes achieved up to 99% accuracy for a binary classification using a naïve classifier. These results are even better than some existing results on k -NN classification. In the future we would like to improve the efficiency of our proposed schemes. We also plan to conduct an experimental study where we consider more than two classes of ground vehicles.

References

1. R. Braunlin, R.M. Jensen, and M.A. Gallo, Acoustic target detection, tracking, classification, and location in a multiple-target environment, Proc. of SPIE Conf. on Peace and Wartime Applications and Technical Issues for Unattended Ground Sensors, vol. 3081, pp. 57-66, Jul. 1997.
2. J. Ding, S.-Y. Cheung, C.-W. Tan, P. Varaiya, *Signal processing of sensor node data for vehicle detection*, Seventh International IEEE Conference on Intelligent Transportation Systems, 2004.
3. M. Duarte and Y.-H. Hu, *Vehicle classification in distributed sensor networks*, *Journal of Parallel and Distributed Computing*, vol. 64 no. 7, pp. 826-838, 2004.
4. I. M. Elfadel and R. W. Pickard, *Gibbs Random Fields, Cooccurences, and Texture Modeling*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, No. 1, Jan. 1994.
5. L. German, *Signal recognition: Both components of the short time Fourier transform vs. power spectral density*, Pattern Analysis and Applications, vol. 6, no. 2, pp. 91-96, 2003.
6. D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, *Detection, Classification, and Tracking of Targets in Distributed Sensor Networks*, IEEE Signal Processing Magazine, 19(2):17-30, March 2002.

7. S. G. Mallat, *A Theory for Multiresolution Signal Decomposition: A Wavelet Representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, July 1989.
8. B. Malhotra, I. Nikolaidis, J. Harms, *Distributed Classification of Acoustic Targets in Wireless Audio-Sensor Networks*, Submitted to the Computer Networks (special issue on Wireless Multimedia Sensor Networks), 2007.
9. G. Succi and T. K. Pedersen, *Acoustic target tracking and target identification—recent results*, Proc. of SPIE Conf. on Unattended Ground Sensor Technologies and Applications, vol. 3713, pp. 10–21, Jul. 1999.
10. H. Wu and M. Siegel and P. Khosla, *Vehicle Sound Signature Recognition by Frequency Vector Principal Component Analysis*, IEEE Transactions on Instrumentation and Measurement, vol. 48, no. 5, Oct. 1999.
11. X. Qin and Y.-H. Yang, *Similarity Measure and Learning with Gray Level Aura Matrices (GLAM) for Texture Image Retrieval*, Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2004.
12. S. Basagni, *Distributed Clustering for Ad Hoc Networks*, Proceedings of the 1999 Intl. Symposium on Parallel Architectures, Algorithms and Networks (SPIN), 1999.
13. M. Weiser, B. Welch, A. Demers, and S. Shenker, *Scheduling for Reduced CPU Energy*, Proceedings of the 1st USENIX Symposium on Operating Systems Design and Implementation., pp. 13–23, November 1994.