Distributed and Efficient Classifiers for Wireless Audio-Sensor Networks

Baljeet Malhotra
University of Alberta, Canada
baljeet@cs.ualberta.ca

Ioannis Nikolaidis
University of Alberta, Canada
yannis@cs.ualberta.ca

Mario A. Nascimento
University of Alberta, Canada
mn@cs.ualberta.ca

***Abstract*—This paper presents schemes to generate effective feature vectors of low dimension, and also presents a cluster–based algorithm, where sensors form clusters on–demand for the sake of running a classification task based on the produced feature vectors. The features generated through our proposed schemes are evaluated using K-Nearest Neighbor (k-NN) and Maximum Likelihood (ML) classifiers. The proposed schemes are effective in terms of classification accuracy, and can even outperform previously proposed approaches, but, in addition, they are also efficient in terms of communication overhead.**

***Index Terms*—Sensors, Acoustic Classification, Features Selection, Data Fusion, Decision Fusion.**

I. Introduction

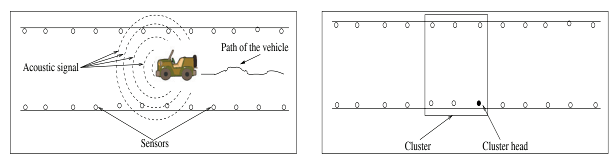
Vehicle tracking on acoustic data is based on the fact that different vehicles produce distinctly different acoustic signals because their engine and propulsion mechanisms are unique [1]. Recently, attention has concentrated on target classification based on acoustic signals using wireless sensor networks [2], [3]. Sensor networks provide redundancy in terms of sensing and processing units. For example, signal measurements can be recorded at multiple sensors. These measurements can then be exploited in a number of ways. Two popular approaches, [2], are data fusion (DAF) and decision fusion (DEF).

Implementing decision or data fusion requires sensors to collaborate efficiently with each other to perform classification. We use a sensor network to perform the classification task within a limited and non-fixed subset of the network, which we call a *cluster*. Clusters can be formed *on-demand* when a signal suggesting the presence of a vehicle is detected by a minimum number of sensors. The purpose of on-demand based clustering is that a static logical structure is not needed to be maintained until an actual classification operation is to be performed.

In this context, we describe in Section II a clustering scheme, which forms the basis of our distributed classification framework. Using the same we can evaluate the benefits of DAF and DEF in various scenarios. We evaluate these approaches and scenarios using K-Nearest Neighbor (*k-NN*) and Maximum Likelihood (*ML*) classifiers, discussed in Section III. These classifiers operate upon suitable feature vectors, which are critical for achieving good classification results. In Section III we describe simple feature extraction schemes that yield low dimensional, yet representative, feature vectors of the captured acoustic signals. In Section IV, we present performance evaluation results in terms of classification accuracy and energy expenditure trade-offs. Finally, Section V summarizes the findings of the paper and outlines our future research goals. (For an extended version of this paper please refer to [4].)

II. On–Demand Clustering

We consider a network of static sensor nodes, and a subset of the network that forms a *cluster*. The classification task is to be performed collectively by the nodes within a cluster. We denote by T the time during which a vehicle is classified. We assume that no more than one vehicle is within the network during T, all sensors have a synchronized global clock, and all sensors gather data every W_t ($< T$) time units.



(a) Vehicle detection. (b) Cluster formation.

Fig. 1. A scenario of vehicle classification in a sensor network.

The problem at hand is to organize the sensors in the form of a cluster, to collectively classify a vehicle that is crossing the sensor network area as shown in Figure 1. When a vehicle enters the network, a number of nodes can detect its presence by measuring its acoustic energy using a detection algorithm similar to the one in [5]. A sensor that detects the vehicle broadcasts a *detection message* (which includes the signal strength of the acoustic signal it recorded) to its neighbors and enters the *active* state. An active node that finds its signal strength greater than its neighboring active nodes, and has received at least N_t detection messages, becomes a *clusterhead*. Otherwise, it waits for W_t before measuring the signal strength again, and, if appropriate, sending a detection message. N_t is an important parameter because, in the interest of classification accuracy, a certain minimum number of active nodes is needed (see section IV). The clusterhead broadcasts a *clusterhead message*, to let all neighboring active nodes know its presence. The clusterhead is free to select only those neighbors that have reported a good signal strength (assuming it selects at least N_t), but in the remaining of this paper we assume that the clusterhead selects all neighboring active nodes.

The clusterhead message includes a *membership list* of the active nodes selected to form the cluster. All active nodes that find themselves on the membership list reply with a *confirmation message* to finalize the cluster formation. Active nodes that receive multiple clusterhead messages select as clusterhead the one that has greater signal strength. Once part of a cluster, an active node does not send further detection messages and also ignores the detection messages from other active nodes until the classification time, T, expires.

Once a cluster is formed, the classification process operates in two steps:

1) The clusterhead prepares and broadcasts a *schedule*. The schedule consists of *task assignments* for all (active) nodes in the cluster. A typical task for an active node is to compute the similarity measure of an unknown sample with respect to the training samples as specified in the schedule. Recall also, that all active nodes in a cluster have their own feature vectors ready to be used for classification purpose.
2) After performing their assigned task, the sensors report back to the clusterhead with their individual results (e.g., a decision or distance measurement computed in response to the first step). After collecting results the clusterhead makes a decision on the class of the unknown vehicle.

When DAF is used, all sensors in a cluster send their feature vectors to their clusterhead. The clusterhead combines all received feature vectors (including one from itself) and executes the classification task using, e.g., k-NN or ML classifiers. When DEF is employed, all sensors in a cluster execute the classification task using their own feature vectors, and make their own decision on the class of the unknown vehicle. The sensors then provide the clusterhead with their decisions. The clusterhead determines the class of the unknown vehicle to be that of the majority of decisions.

III. Classification Techniques and Features Selection

If we assume d to be the length of the feature vectors, and l to be the number of training samples in each of the c classes then the number of computations performed by a sensor to classify an unknown sample is proportional to $d \times l \times c$ and d^2 when employing k-NN and ML respectively. Selection of d is thus critical as we would like to reduce the computational load from sensors, of-course, without compromising on the quality of solution. In addition, the size of a feature vector is also crucial for communication costs in sensor networks (recall DAF). That is, a larger d causes a higher rate of energy consumption for the communication among the sensors.

Other researches have used different techniques to extract feature vectors from acoustic signatures. Duarte *et. al.* in [3], first choose 100 FFT points from the fast Fourier transform of 512 data points sampled at a rate of 4.960 kHZ. Then, they average the 100 FFT points by pairing consecutive points to get a 50 dimensional feature vector. Brooks *et. al.* [2] also used a 50-dimensional FFT feature vectors extracted from the time series data. Wang *et. al.* [6] used PCA to choose the 15 largest eigenvalues to form the eigenspace for their training and test data. Unfortunately, selecting the first few principal components provides only a measure of statistical significance without guaranteeing to yield the best subset of features. The reason is that PCA finds feature combinations that model the variance of a data set, but these may not be the same features that separate the classes.

These pre-existing methods are either simply not meeting the demands of energy conservation in sensor networks [3], or they are generic in nature, e.g., PCA, and computationally expensive, while not even yielding the best results. More importantly, however, they do not address the issue of relating dimensionality of the feature vectors with the objective of energy conservation. What is required is to satisfy two competing demands: that of creating feature vectors that are low on dimensions, and that of being able to produce good classification results. Towards that goal we present two schemes for feature extraction from the acoustic signatures of vehicles.

A vehicle sound is a stochastic signal and, in practice, the sound of a moving vehicle observed over a short period of time can be treated as a stationary signal [7]. In our case we use the signal's duration to be 51.6 ms, *i.e.*, 256 data points sampled at a frequency of 4.960 kHz. In our study we considered power spectral density (PSD) based features. This feature is generated by taking PSD estimates of 256 data points yielding a linear vector of 128 PSD points with a resolution of 38.75 Hz. In the rest of the discussion a PSD point is also called a band of frequencies, or simply a dimension, because a PSD point represents a collection of consecutive frequencies.

Our proposed schemes start by considering all 128 dimensions, and subsequently pruning many of them. The basis of our proposed schemes, and the first pruning criteria is that most of the power in a vehicle's sound lies in the lower frequencies. To create a feature vector our schemes start by choosing only those dimensions that correspond to frequencies that have the maximum power as reported by the samples of the corresponding training class.

Specifically, let f_i^j be a frequency band that has the maximum power as reported by the sample j in the class i. Let S_i be the set of all f_i^j's reported by all samples j in the class i. Note that $|S_i| \leq l$, i.e., some samples in class i may report on a common dimension. This particular situation is favorable for producing feature vectors that are low on dimension, and yet be effective. Our intuition is that a dimension that has been reported by a large number of samples in a class is more suitable to characterize that particular class than a dimension which has not. Following this intuition, we rearrange S_i. First, we count the number of times each unique dimension has appeared in S_i to obtain their rank. Then, we order the dimensions in S_i in decreasing order of rank. After rearranging, S_i, we further prune some more dimensions by selecting a percentage, ρ, of top ranked dimensions from S_i to constitute another set, S_i^ρ. By this pruning criteria, we eliminate those dimensions that are less frequent in the training class. This process is repeated, in order to derive the S_i^ρ set for each class i = 1,2...c.

We propose two schemes to select elements (dimensions) from sets S_i^ρ to create the feature vectors, which will be stored in the sensor nodes. In the first approach the sets S_i^ρ are combined to obtain the set $S = \bigcup_i S_i^\rho$. We name this approach an independent feature selection (IFS) scheme. In the second approach, features are selected by considering only those dimensions from sets S_i^ρ that are common to *all* classes, that is, $S = \bigcap S_i^\rho \; \forall i$. We name this latter approach global feature selection (GFS), as dimensions are common to all training classes. A potential problem that might occur in the GFS scheme is that the final set S may remain empty if there is no common dimension among the sets S_i^ρ. Since we can control the size S_i^ρ by setting an appropriate value for ρ, we can handle this exception by increasing the value of ρ. If the set S remains empty even for $\rho = 1.0$, the first element from all sets S_i^ρ is chosen to be inserted into the final set, S.

IFS and GFS feature vectors can be obtained in advance from a training set. The feature vectors of the training samples are then uploaded to the sensors before their deployment. After deployment, the sensors can extract 128 PSD points from the time series of the sampled acoustic signal of an unknown vehicles, and can directly fetch IFS/GFS feature vectors of the unknown sample using the selected dimensions learned from the training phase.

IV. Experimental Study

In this section we present the results of our experimental study where we evaluate the performance of our distributed classification schemes as well as the merits of feature vectors generated through our proposed feature extraction schemes. We are mainly interested in comparing accuracy results with the results from already existing studies. With that in mind, we chose an acoustic dataset that has been used elsewhere for similar studies. The dataset we consider was generated during the 3rd SensIT situational experiment (SITEX02), organized by DARPA/IXOs SensIT program. We call this the *SensIT dataset* (http://www.ece.wisc.edu/~sensit). The dataset is standardized to remove shifting and scaling factors by using the *normal form* [8] of the original time series data.

One of the challenges in our experimental study was to simulate a distributed environment of a sensor network. The signal of an unknown vehicle captured by a sensor may be different from the signal from the same vehicle captured by another sensor at approximately the same time. This is due to the placement of the sensors. In order to create multiple synthetic copies of an acoustic signal, to represent what different sensors would have acquired, we adopted the following procedure: For each acoustic signal selected from our dataset to play the role of the unknown sample, we create multiple copies of the same, attenuating the original signal based on the distance of the sensors from the moving vehicle. Then, we introduced time difference of arrival lags for the sensors based on their relative position with respect to the moving target. We also added white noise for each of the sensor's signal. Finally, we standardize the (attenuated and noised) synthetic signal, by applying the normal form. This procedure is repeated for every testing signal in our dataset. For the sake of simplicity we assume the environment to be such that reverberation and Doppler effect can be safely ignored as negligible.

We consider two performance metrics: (i) classification accuracy, and (ii) energy expenditure. A sample is considered correctly classified if the true class is predicted. In the k-NN method similarity between any two samples is computed using $L1$ distance metric. Energy expenditure is computed based on the number of bits transmitted by a sensor. We assume the same radio model as in [9], according to which a sensor spends $50 + 0.1 \times R^3 \; nJ/bit$ of energy to send one bit at R distance. We note that the cost of assembling the cluster is the same for both data and decision fusion approaches and therefore it has no impact on distinguishing which scheme is more energy efficient

Before implementing the proposed classification schemes we studied the relationship between the size of the training classes and the size of the feature vectors generated through our proposed IFS/GFS schemes. Our findings, described in detail elsewhere [4], can be be summarized as follows.

With a particular training class size, when we allow to select more of the top ranked dimensions (by increasing ρ), the size of feature vector naturally increases for both of the schemes. However, after one point, namely when the training class size increases from 108 samples/class to 180 samples/class, the size of the feature vector does not change much. This behavior happens because when the size of the training class is sufficient the consensus among the samples of the training classes is high, hence, adding more samples into the training classes does not affect the consensus much. As expected, with a similar setting of parameters for IFS and GFS schemes, GFS produced feature vectors of smaller size.

We used both IFS and GFS to generate feature vectors for the k-NN and ML classifiers. k-NN and ML classifiers obtained different classification accuracies with various settings of IFS and GFS schemes. The best classification accuracies are reported here and compared with the previously achieved best accuracies on the same dataset. In particular the results reported in this section were obtained with a training class of 45 samples/class and ρ value of 0.3 for GFS scheme in the k-NN classifier, and 63 samples/class and ρ value of 0.5 for IFS scheme in the ML classifier. With this setting, the average size of the feature vectors was found to be 8, which is almost 1/6th the size of the features vectors used in [2], [3]. In general, selecting 30-50% of the top ranked dimensions (i.e. a ρ value in the range [0.3,0.5]) produced the best results.

Table I summarizes and compares our classification results with the results from the previous studies on the same dataset using various decision and data fusion approaches.

In the study of Brooks *et al.* [2], various scenarios of data fusion and decision fusion were considered using single and

TABLE I
COMPARISON OF CLASSIFICATION ACCURACIES.

Results from	Accuracy (%)			
	k-NN		ML	
	DEF	DAF	DEF	DAF
This paper	**77.89**	77.63	**89.46**	**89.20**
Brooks [2]	–	–	77.90	81.30
Duarte [3]	69.36	–	68.95	–
Wang [6]	–	**84.68**	–	–

multiple sensors. We were mainly interested in evaluating the impact of feature vectors in various settings of data and decision fusion approaches using the well known classifiers and a cluster logical structure. Their compared results presented in Table I are from the acoustic modality with decision fusion from multiple sensors. The results from the study of Wang *et. al.* [6] on acoustic data compared in Table I are based on a data fusion approach for which they modified the multi-resolution integration (MRI) algorithm originally proposed by Prasad *et. al* in [10]. Duarte *et. al.* [3], used acoustic as well as seismic modality in their study. Their results presented in Table I are based on the local classification with decision fusion using the acoustic modality.

Through Table I, we have compared the best results from the studies mentioned previously with the best of our results obtained with various settings of data and decision fusion schemes proposed here. In particular our decision fusion approach, DEF, using GFS feature vectors produced the best classification accuracy of 89.46% as compared to all other results presented for the ML classifier.

Clustering has significant impact on the results obtained through data and decision fusion approaches. As shown in Table II accuracy for both approaches improves slightly as we increased the cluster size. The reason for improved accuracy is that increasing the number of sensors in the cluster increases the probability of making a correct prediction. However, after a sufficient number of sensors are available within the cluster, 20 in the case of our experiments, adding more sensors did not seem to improve the accuracy.

TABLE II
ACCURACY AND COST FOR DIFFERENT CLUSTER SIZES.

Cluster Size	Accuracy (%)				Cost (μJ/sensor)			
	k-NN		ML		k-NN		ML	
	DEF	DAF	DEF	DAF	DEF	DAF	DEF	DAF
3	70	68	88	83	17	21	17	21
5	70	71	89	89	19	25	19	24
10	72	73	88	89	29	38	26	37
20	74	77	89	89	67	83	67	82
40	78	76	89	88	209	240	210	239

Clustering has much more impact on the energy expenditures than on the accuracy. As can be seen in Table II, when the number of sensors in the single cluster increases from 3 to 40 sensors the energy expenditures increased substantially. The reason is that the increased number of sensors causes more communication exchanges between the cluster members. In particular, the data fusion approach, DAF, incurred more cost due to the transmission of feature vectors by the sensors.

V. CONCLUSIONS AND FUTURE DIRECTIONS

Classifying audio signals is an important application in wireless sensor networks. We proposed two distributed classification schemes, which take into account the inherently distributed nature of the problem and produce competitive classification results. Our proposed feature extraction schemes are generic, and may find applications in other areas where feature selection is a difficult task due to the high dimensionality of feature vectors. One limitation of our proposed schemes that requires further work is finding the right size of training classes, and setting an appropriate value for ρ.

In the context of our application domain, another promising venue for further work is to allow the classification process to be a continuous along the vehicle's path. Towards that goal we are currently investigating how to perform the classification task at several different clusters, and how clusters should communicate with each other in order to improve the classification accuracy, while energy efficient.

ACKNOWLEDGMENT

This research is partially supported by NSERC.

REFERENCES

[1] 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, 1999.

[2] R. R. Brooks, P. Ramanathan, and A. M. Sayeed, "Distributed target classification and tracking in sensor networks," *Proc. of the IEEE*, vol. 91, no. 8, pp. 1163–1171, 2003.

[3] M. Duarte and Y. H. Hu, "Vehicle classification in distributed sensor networks," *J. of Parallel and Distributed Computing*, vol. 64, no. 7, pp. 826–838, 2004.

[4] B. Malhotra, I. Nikolaidis, and M. A. Nascimento, "Distributed and efficient classifiers for wireless audio-sensor networks," Dept. of Computing Science, Univ. of Alberta, Tech. Rep. TR08-02, 2008. [Online]. Available: http://www.cs.ualberta.ca/TechReports/2008/TR08-02/TR08-02.pdf

[5] J. Ding, S. Y. Cheung, C. W. Tan, and P. Varaiya, "Signal processing of sensor node data for vehicle detection," *Proc. of the 7th Intl. IEEE Conf. on Intelligent Transportation Systems*, pp. 70–75, 2004.

[6] X. Wang and H. Qi, "Acoustic target classification using distributed sensor arrays," *Proc. of the IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 4186–4189, 2002.

[7] H. Wu, M. Siegel, and P. Khosla, "Vehicle sound signature recognition by frequency vector principal component analysis," *IEEE Trans. on Instrumentation and Measurement*, vol. 48, no. 5, pp. 1005–1009, 1999.

[8] D. Goldin and P. Kanellakis, "On similarity queries for time-series data: Constraint specification and implementations," *Proc. of the 1st Intl. Conf. on the Principles and Practice of Constraint Programming*, pp. 137–153, 1995.

[9] W. H. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," *Proc. of the 33rd Annual Hawaii Intl. Conf. on System Sciences*, vol. 2, pp. 1–10, 2000.

[10] L. Prasad, S. S. Iyengar, and R. L. Rao, "Fault-tolerant sensor integration using multiresolution decomposition," *Physical Review E*, vol. 49, no. 4, pp. 3452–3461, 1994.