# Object Detection Using Acoustic Sensors for Surveillance Applications

Selver Ezgi Küçükbay[1], Mustafa Sert[2] and Adnan Yazıcı[3]

***Abstract*—**Nowadays, acoustic sensors for environmental sound classification are becoming very important for surveillance applications. The use of the acoustic sensors for gathering and processing real time audio data are essential because they are stealth and requires low energy. In this study; we design and implement a wireless acoustic sensor network for object detection. In this context, we propose an acoustic wireless sensor hardware design to gather sounds for surveillance applications. We extract the features from each sound clip and classify them to detect objects. We construct a dataset for detection of human, animal and vehicle classes. We have done a number of experiments on this dataset and evaluate the results. The overall accuracy of our solution, based on our experiments, is 88.3%.

***Keywords*—**acoustic sensors, MFCC, object detection, surveillance applications, Raspberry Pi, SVM.

## I. Introduction

In the recent years, with the rapidly growing of Internet of Things (IoT), network-based sensor applications have gained popularity. This concept is used in many areas such as health, military, smart home systems and surveillance applications. Especially, sound related studies have an important role in these applications [1-5]. Using acoustic sensors, we can collect data in real time for analysis, object detection and classification for the purpose of surveillance. In object detection, acoustic sensors are most widely used in the literature since they don't need image and they are energy efficient.

In the literature, there are a number of studies related to acoustic sensors and sound classification. In [1], they collect vehicle sounds with acoustic sensors and classify them using Gaussian Mixture Model-GMM. The acoustic signatures of the vehicles vary due to different factors such as speed, accelerations, gear states and engine speeds. In many cases, it depends on the position of the sensor and the road conditions. This problem becomes more important for the sensors that are connected to each other over the wireless network rather than size and energy constrains. The dataset used in that study contains 146-minute vehicle sounds. The vehicles are classified as big, small, medium sized vehicles and motorcycle. In the dataset contains 1123 "light" and 65 "heavy" vehicle sounds. As a result of binary classification, with the GMM, they achieved 7% error rate.

In [2], the authors use acoustic sound signals for classifying vehicle-terrain interaction in their study. Acoustic data are collected by microphones on mobile robots in different terrains and objects in the outdoor environment. The data are labeled and trained with supervised multi class classifier. The features from acoustic signals are extracted from both time domain and frequency domain. For the time domain features, they use 3 different feature vectors such as zero crossing rate (ZCR), short time energy (STE) and energy entropy. They also extract various features from frequency domain. Since they study multi-class classification, they use support vector machine (SVM) classifier with one-versus-one approach. Dataset that they collected from sensors contains 6 different outdoor environments and each environment is labeled as a class. The classes fall in to two main categories and three classes for each category. The main categories are benign terrain and hazardous terrain. The classes under benign terrain are: driving over grass, driving over pavement and driving over gravel road. Classes under hazardous terrain are: splashing in water, hitting hard objects and wheels losing traction in slippery terrain. The experimental results give 92% accuracy. In [4], they focus on a single sensor problem and classify the types of the moving vehicles. For vehicle classification, the hybrid dictionary learning method is proposed by the authors. They collect vehicle sounds from sensor nodes and extract mel frequency cepstral coefficients (MFCC) features from the clips. According to their experimental results, they yield 88.11% accuracy. In [5] , the feature extraction methods are studied for object detection with passive acoustic sensors deployed in suburban environments. The sound classes are gun, vehicle and pedestrians. The gun class contains muzzle blast, the vehicle includes running piston engine and pedestrians consist of several footsteps. Based on their study, the feature extraction methods related to physical nature of the objects and since the location of the object is required for one of the extraction methods, a time-based location method is also developed. The dataset are collected in real time and suggested methods are evaluated on experimental acoustic sensor nodes. In [9], an acoustic sensor based system is proposed for human activity recognition. The proposed system classifies human activities using acoustic information. The dataset in the study contains synthetic sounds. The human activities are drinking soup,

Selver Ezgi Küçükbay[1] is with the Başkent University, Department of Computer Engineering, and Middle East Technical University, Department of Computer Engineering, Ankara, Turkey

Mustafa Sert[2] is with the Başkent University, Department of Computer Engineering, Ankara, Turkey

Adnan Yazıcı[3] is with the Nazarbayev University, School of Science and Technology, Department of Computer Science, Astana, Kazakhstan

eating rice, putting down metal spoon, putting down bowl (porcelain, plastic), drinking tea (with a saucer, without a saucer, hot tea). Their proposed system is evaluated using different classifiers such as the J48 decision tree, Naive Bayes and SVM.

In this paper, we specifically study object detection in wireless surveillance networks using acoustic sensors. We determine the classes, animal, human and vehicle, which are important in surveillance applications. We also record different actions of human, various animal or vehicle sounds to increase robustness of the system that we developed. For that, we use a proper circuit design for acoustic sensors and collect the data using these sensors for surveillance applications. We concentrate on the signatures of acoustic signals so that we extract the features (i.e., MFCC) from the acoustic data and then we classify the sounds using the SVM classifier on the sensor nodes including Raspberry Pi. We have done a number of experiments on this dataset that we gathered and evaluated the test results.

This paper is organized as follows: In Section 2, designed system, feature extraction methods and classifier details are presented. In Section 3, experimental results are given and finally in Section 4, the results of our experiments are reported.

## II. Materials and Methods

In this study, we design a wireless sensor node which are using acoustic sensors to collect real time acoustic sounds. The feature extraction is performed on real time data and classification process is applied based on feature sets. The main steps of the study are presented in Fig. 1.

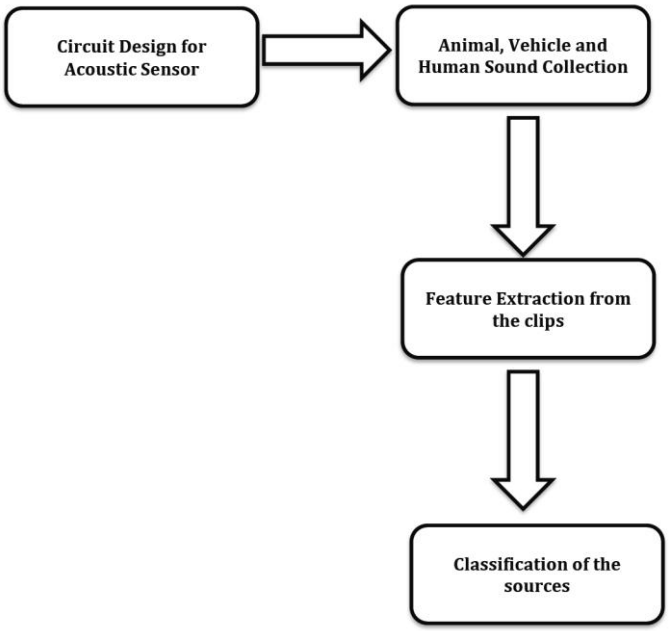


Fig. 1 General design of the system

TABLE I
Acoustic Data Set

| Class Name | Number of the Samples |
|---|---|
| Animal | 265 |
| Human | 240 |
| Vehicle | 219 |

### A. Acoustic Sensor Data Set

We define our acoustic sound classes as human, animal and vehicle for the surveillance application. For these classes, we collect the samples from the acoustic sensor nodes. The collected data are gathered from different sound sources. For vehicle class, bus, car and motorcycle sounds are recorded. For human class, crowd sound, footstep, speech and laugh sounds of female and male are taken. The animal class contains cat, bird, dog sounds. In addition to data collection, we extract the features and design a classifier in order to classify these data. Table I shows the class names and number of the samples of each class.

### B. Circuit Design for Acoustic Sensor

The acoustic sensors that are used in this research produce analog output. We use Raspberry Pi[1] in order to collect data from sensors. Since the Raspberry Pi can only read digital signals, output signals from sensors in the study expressed as numeric values using analog-to-digital converter (ADC). We used 12 bit 1 channel MCP3201 ADC [6] for our circuit design. The MCP3201 ADC is connected to Raspberry Pi using the pins on the data sheet. Reading data is done by using Serial Peripheral Interface (SPI). SPI is a standard, which provides synchronous serial communication for short distance communication, especially in embedded systems. The communication between Raspberry Pi and SPI is done by data bit transfer. Fig. 2 shows the circuit design between sensor, MCP3201 ADC and Raspberry Pi connections.

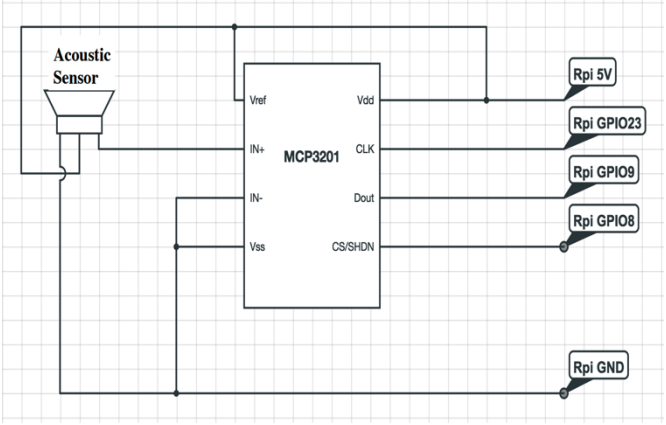


Fig. 2 Circuit design for the acoustic sensor

### C. Feature Extraction

We choose MFCC feature extraction method [7] to extract the characteristic of the sounds. MFCC is widely used in

[1] www.raspberrypi.org

literature due to their success in the speech recognition applications.

Basically, the MFCC is modeled as a human perception. In the MFCC, signal is processed in frames rather than as a whole. In this study, we choose Hamming windowing function for framing the signal. In the framing method, since we lost the information from the endpoints, this problem is solved by overlapping the windows and determining the hop sizes. In this study, we choose 30 ms as a window size and 10 ms as a hop size. These window, hop sizes and coefficients define the dimension of the feature vector. The feature vector (F) is represented as a (nxm) matrix. n is the number of analyzed window number and m is the coefficient of each analyzed windows. Each sound clip in the dataset represented as a nxm matrix in our study. We used 13-coefficient MFCC. In order to decide the classes on clip based and reduce the complexity, we convert the nxm matrix to 1xm matrix. For converting the matrix, we are calculating column-based arithmetic mean [12,13]. As a result, for each clip, we gather 1x13 feature vector. Model training in the classifier design and tests are done by using these vectors.

*D.Classifier Design*

As a classifier design, SVM is selected since it gives reasonably good results in speech and pattern recognition applications [7,10-11]. LIBSVM library is utilized for implementing SVM [8]. SVM is a binary class classification method. In the study, we have three different classes: human, vehicle and animal. Since this is a multi class classification problem, we use one-versus-all method to modify SVM for multi class problem. According to this method, a particular model is created for each class, the samples that are belonging this class are identified and the rest of the samples are distinguished according the model. For our research, we have 3 different models for animal, vehicle and human.

In the test phase, the test samples assigned the classes according to maximum probability. In order to evaluate the classifier design, we use 5 fold cross validation. According to this, data set is split in to 5 parts. One part is reserved for the test and the remaining four parts are used in model training. The part that is used for the test set in the previous phase is included in the model training at a later stage and one of the parts used in model training is allocated for the test. By repeating this process five times, each set used in both training and test for more reliable results.

## III. Experimental Results

Evaluations are reported as precision, recall and F-measure in the study. The gathered information from the experiments can be expressed by confusion matrix. According to this matrix the results are labeled as true positive (TP), false negative (FN), false positive (FP) and true negative (TN).

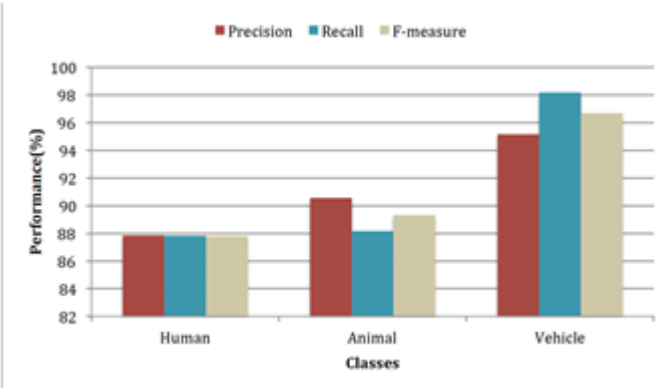


Fig 3. Class-based results

We yield 88,3% performance for three different classes in our study. Fig. 3 shows the results in terms of precision, recall and F-measure.

Table II shows the fold-4 confusion matrix as an example. According to this matrix, a sample actually belongs the animal class is predicted as a human. The sample labeled with a animal class is predicted as a human by the system. In this case, in the presented study, human and animal classes are mixed by the system. Thus, although the performance for the vehicle class is very good, the average accuracy of the system falls a little due to animal-human confusion. To avoid this problem, more data need be collected for human and animal classes. Unfortunately, collecting data using acoustic sensors is tedious and time consuming, and hard to set a real-life test environment. If more samples were collected, the system would learn the patterns much better. Thus, when a new sample comes for testing, it can be easily add to the classifier.

TABLE II
FOLD-4 CONFUSION MATRIX

| | | *Predicted* | | |
|---|---|---|---|---|
| | | **Animal** | **Vehicle** | **Human** |
| *Actual* | **Animal** | **44** | 3 | 6 |
| | **Vehicle** | 0 | **44** | 0 |
| | **Human** | 9 | 2 | **37** |

According to the results in Figure 3, selected feature extraction and classifier methods give above 80% accuracy in all categories. The proposed system achieves reasonably good results for human, animal and vehicle detection in wireless surveillance networks.

## IV. Conclusion

In this study, we aim to introduce a wireless sensor node design for object detection using acoustic sensors. Animal, human and vehicle sounds are collected using the circuit (sensor nodes) and are classified these acoustic data using the SVM machine learning algorithm. MFCC feature extraction method is applied to collected data. The features are classified using SVM classifier. Average accuracy of SVM classifier for this surveillance application is 88.3%. For more robust system, deep-learning algorithms may be used. However, we must note that using deep learning requires much bigger data; otherwise using deep learning approach may not give better performance. Therefore, we have been working on applying the deep learning approach along with transfer learning with small dataset aiming possibly a better performance. As another important future study is to expand the dataset and make it publicly available for the other researchers in the field to use it for evaluating their systems.

## Acknowledgment

This project is supported by TUBITAK (Project Number: 114R082).

## References

[1] Burhan F. Necioglu, Carol T. Christou, E. B. George, Garry M. Jacyna, "Vehicle acoustic classification in netted sensor systems using Gaussian mixture models", *Proc. SPIE 5809, Signal Processing, Sensor Fusion, and Target Recognition XIV*, 409 (May 31, 2005).

[2] J. Libby, A. Stentz, "Using sound to classify vehicle-terrain interactions in outdoor environments", *IEEE International Congerence on Robotics and Automation (ICRA)*, 2012.

[3] J. Lee, H. J. Kim, "Acoustic classification and tracking of multiple targets using wireless sensor networks", *15th International Conference on Control, Automation and Systems (ICCAS 2015)*, 2015. https://doi.org/10.1109/ICCAS.2015.7364947

[4] S. Guo, R. Wang, B. Liu, Q. Wei, Y. Li, "Vehicle classification in acoustic sensor networks based on hybrid dictionary learning", *IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, 2016. https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2016.147

[5] T. H. Groot, E. Woudenberg, A. G. Yarovoy, "Urban objects classification with an experimental acoustic sensor network", *IEEE Sensors Journal*, Vol. 15, No. 5, 2015. https://doi.org/10.1109/JSEN.2014.2387573

[6] MCP3201 2.7V 12-Bit A/D Converter with SPI Serial Interface, Microchip Technology Inc, 2007.

[7] L. Chen, S. Gunduz, M. T. Ozsu, "Mixed type audio classification with support vector machine", *IEEE International Conference on Multimedia and Expo*, July 2006, pp. 781–784. https://doi.org/10.1109/ICME.2006.262954

[8] C.-C. Chang, C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol*., vol. 2, no. 3, pp. 27:1–27:27, May 2011.

[9] J. M. Sim, Y. Lee and O. Kwon, "Acoustic sensor based recognition of human activity in everyday life for smart home services", *International Journal of Distributed Sensor Networks*, Vol.11, Issue 9, 2015..

[10] J. C. Wang, J. F. Wang K. W. He, C. S. Hsu, "Environmental sound classification using hybrid svm/knn classifier and mpeg-7 audio low-level descriptor", *International Joint Conference on Neural Networks*, 2006 https://doi.org/10.1155/2015/679123.

[11] Ç. Okuyucu, M. Sert, A. Yazıcı, "Audio Feature and Classifier Analysis for Efficient Recognition of Environmental Sounds," *IEEE International Symposium on Multimedia (ISM)*, s.125–132, 2013. https://doi.org/10.1109/ISM.2013.29

[12] S.E. Küçükbay and M. Sert,"Audio Event Detection Using Adaptive Feature Extraction Scheme", *The Seventh International Conferences on Advances in Multimedia*, Barcelona, Spain, pp. 44-49, 2015.

[13] S.E. Küçükbay and M. Sert, "Audio-based event detection in office live environments using optimized MFCC-SVM approach", *IEEE International Conference on Semantic Computing (ICSC'2015)*, Anaheim, CA, USA, pp.475-480, 2015.

**Selver Ezgi Küçükbay** received her BSc and MSc degrees in Computer Engineering from Başkent University, in 2012 and 2015, respectively. She is currently a PhD candidate at Computer Engineering Department in Middle East Technical University. She is working as a research engineer at Başkent University, Ankara, TURKEY. Her research area is about pattern recognition, machine learning, multimedia systems.

**Mustafa Sert** received the MSc and PhD degrees in computer science from Gazi University, Ankara, Turkey, in 2001 and 2006, respectively. He has been an assistant professor of computer engineering at the Baskent University since 2006. He has research interests in theory and applications of audio signal processing, machine learning, pattern recognition, and multimedia databases. He mainly focuses on semantic content extraction from audio and video data, audio scene recognition, video concept detection, multimodality, and content modeling for multimedia search and retrieval. He serves in technical reviewing and organization committees of several international conferences including IEEE IRC 2017,VLDB 2012, FUZZ-IEEE 2015, IEEE ISM (2008-2009, 2017), FQAS 2009, and IEEE BigMM 2016. He also serves as a reviewer of the IEEE&ACM TASLP, IEEE SPL, Springer MTAP, Springer SIVP, and IEEE TFS. Dr. Sert is a senior member of the IEEE and the IEEE Computer Society.

**Adnan Yazıcı** received his Ph.D. degree in computer science from the Department of EECS, Tulane University, LA, USA, in 1991. Before joining Nazarbayev University as the Chair of Computer Science, he had been a Full Professor and the Chair of the Department of Computer Engineering, Middle East Technical University, Ankara, Turkey, where he had been also the Director of the Multimedia Database Laboratory. He has published over 200 international technical papers and co-authored/edited three books entitled Fuzzy Database Modeling (Springer), Fuzzy Logic in its 50th Year: New Developments, Directions and Challenges (Springer), and Uncertainty Approaches for Spatial Data Modeling and Processing: A Decision Support Perspective (Springer).

He was a recipient of the IBM Faculty Award in 2011 and the Parlar Foundation's Young Investigator Award in 2001. He was the Conference Co-Chair of the 23rd IEEE International Conference on Data Engineering in 2007, the 38th Very Large Data Bases in 2012, and the 23rd IEEE International Conference on Fuzzy Systems in 2015. He is currently an Associate Editor of the IEEE Transactions on Fuzzy Systems and a member of the Fuzzy Systems Technical Committee of the IEEE Computational Intelligence Society.