

Identifying Microphone from Noisy Recordings by Using Representative Instance One Class-Classification Approach

Huy Quan Vu, Shaowu Liu

School of Information Technology

Deakin University, 221 Burwood Highway, Vic 3125, Australia

Email: {quan, swliu}@tulip.org.au

Xinghua Yang

High-performance network laboratory, Institute of Acoustics

Chinese Academy of Sciences, Beijing 100190, P.R. China

Email: yangxh@hpl.ac.cn

Zhi Li, Yongli Ren

School of Information Technology

Deakin University, 221 Burwood Highway, Vic 3125, Australia

Email: zhilimailbox@yahoo.com.au, yongli@tulip.org.au

Abstract—Rapid growth of technical developments has created huge challenges for microphone forensics - a sub-category of audio forensic science, because of the availability of numerous digital recording devices and massive amount of recording data. Demand for fast and efficient methods to assure integrity and authenticity of information is becoming more and more important in criminal investigation nowadays. Machine learning has emerged as an important technique to support audio analysis processes of microphone forensic practitioners. However, its application to real life situations using supervised learning is still facing great challenges due to expensiveness in collecting data and updating system. In this paper, we introduce a new machine learning approach which is called *One-class Classification* (OCC) to be applied to microphone forensics; we demonstrate its capability on a corpus of audio samples collected from several microphones. In addition, we propose a representative instance classification framework (RICF) that can effectively improve performance of OCC algorithms for recording signal with noise. Experiment results and analysis indicate that OCC has the potential to benefit microphone forensic practitioners in developing new tools and techniques for effective and efficient analysis.

Index Terms—Machine Learning, Data Mining, Audio Forensics, Microphone Forensics, One-Class Classification

I. INTRODUCTION

Microphone forensics is a sub-category of audio forensic science, which aims to establish whether an obtained audio recording is original, or to verify whether it was made on a given recorder. The determination of microphone model of arbitrary recording can help assure the

This paper is based on "Microphone Identification using One Class-Classification Approach" by H.Q. Vu, S. Liu, Z. Li, and G. Li, which appeared in the Proceedings of the 2nd Workshop on Applications and Techniques in Information Security (ATIS), Melbourne, Australia, November 2011

actual ownership of that recording in the case of multiple claims of ownership, and thus provides a valuable mechanism to resolve copyright disputes. Rapid growth of technical developments in the past decade has created huge challenges with availability of numerous digital recording devices and massive amount of recording data. These digital media make undetected forgeries and manipulations easy, and might thereby encourage criminals. Demand for fast and efficient methods to assure integrity and authenticity of information is becoming more and more important in criminal investigation nowadays.

Recently, machine learning has emerged as an important method of automated audio analysis processes which provides supportive tools for microphone forensic practitioners. A first attempt of practical evaluation on recording devices and environment classification was performed by Kraetzer et al. in 2007 [1]. They incorporated the *K-means* and *Naive Bayes* as classifiers, and evaluated their classification capability on a set of audio *steganalysis* features. Later on, they proposed an *Unweighted Fusion* framework using a *Decision Tree* and *Linear Logistic Regression* models that achieved higher performance on microphone detection task [2]. In 2009, another attempt using supervised machine learning methods (*Simple Logistic*, *J48 decision tree*, *K-nearest neighbor*, *Support Vector Machine—SVM*) was performed by Buchholz et al. with *Fourier* coefficient histogram extracted from near-silence segments of the recording as the feature vectors [3]. Similar approach was utilized by Garcia-Romero and Espy-Wilson [4] with SVM classifier to assess the performance of *linear-cepstral coefficients* and *mel-scaled cepstral coefficients* (MFCCs) as audio features. The audio samples were obtained from two classes of acquisition devices, land-line telephone handsets and microphones.

Recently, Kraetzer et al. [5] designed a context model to devise empirical investigation to identify suitable classification algorithm and appropriate audio features as a supportive guidance for microphone forensic researchers, the experiment involved 74 supervised classification algorithms and 8 clusters implemented in Weka and 590 intra-frame audio features.

Although, great effort has been spent on this task of microphone forensics, only a limited number of approaches were found. Current techniques using supervised machine learning are still facing great challenges due to expensiveness in collecting data and updating system. Research results are still far from standard for real life application. In this paper, we introduce a relatively new approach of machine learning - *One-class classification (OCC)*, into microphone forensics, which has potential to benefit microphone forensic practitioners in developing new tools and techniques for more efficient analysis. The advantage of OCC is that it builds one classification model for each microphone and based on only data samples of target microphone. Therefore, there is no need to collect all available microphone data samples, and frequently retrain the existing models when new microphone becomes available as supervised learning approach. We also consider the fact that audio recordings usually contain noise, which has negative influence to classification performance. None of existing works have yet tackled this challenges in microphone forensic field. In this study, we propose a representative instance classification framework (RICF) to improve OCC performance on microphone forensic tasks for noisy recording signal. The idea of RICH is to reduce impact of noise on training data set by sampling instances that best describe microphone characterizes from given audio records. And the prediction for a given audio record is based on its probability of belongs to a particular target microphone. The detail analysis of OCC approach and RICH will be presented in later sections.

Having introduced the research motivation, section II provides a critical analysis of supervised learning problems in microphone forensics, following by formulation of research objective for this work. In section III, we describe our audio data collection and select a set of audio features, following by a preliminary analysis of the data set. Section IV is devoted to present several *OCC* techniques as well as representative instances classification framework which are used for our experiments. In section V, we present analysis of empirical results on audio signal recorded from different scenarios. Section VI concludes the paper with a summary and offers practical implication as well as future research suggestions.

II. PROBLEM DEFINITION AND RESEARCH OBJECTIVE

Traditionally, microphone identification is considered as an *n*-class supervised learning problem, where an audio sample is classified into one of *n*-classes of microphone models in the training data. A single model is trained on a data set consisting of audio samples for *n* microphone models. For microphone verification,

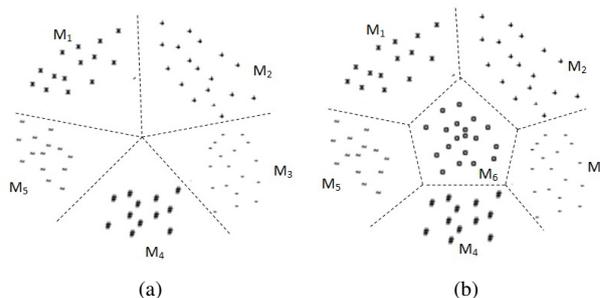


Figure 1. Supervised Machine Learning Approach

the task can be considered as a binary classification problem which aims to determine whether the audio sample is really recorded by a particular microphone models or not. The training data contains of two classes label: *positive* (audio samples of claimed microphone) and *negative* (audio samples of all other microphones), each microphone requires a unique model to verify its identity. Agreed that binary classification is usually easier than multi-class classification, microphone verification would come out to be simpler than microphone identification. However, real world application of microphone forensics using supervised learning approach is a very challenging problem because of their *open* nature. It is impractical, if not impossible, to construct a complete database with audio samples recorded from all available microphones models in the world as training data set. In addition, when new models are being produced continuously, the entire supervised classification model needs to be re-trained frequently, and hence makes supervised learning an expensive approach. This aspect of supervised learning approach is demonstrated in Fig. 1.

Fig. 1(a) shows a decision boundary on a data set containing 5 microphones ($M_1 - M_5$) constructed by supervised machine learning approach, data extracted from different microphones is represented by different symbols. This decision boundary can be used to classify a new data samples into one out of 5 classes corresponding to 5 microphone models. However, when a new microphone becomes available (M_6), the decision boundary need to be retrained in order to correctly classify the microphone models as shown in Fig. 1(b), For these reasons, it is worth to analyzing this task under a different approach that should be more practical, cost effective and easier for system maintenance.

Due to recent advances in machine learning, an approach has emerged to be more suitable for applications of microphone forensics, which is referred to as *One-class classification*. It can be simple to obtain audio samples from a particular microphone model as *positive* (target) class but impossible to collect from all other available microphones as *negative* (outlier) class. The idea behind OCC approach is to design a classifier so that only the target class is characterized, and consequently can distinguish it from all counter-examples from outlier classes. Note that, OCC approach characterizes only the

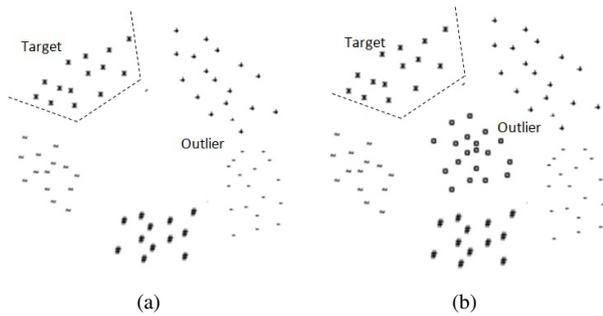


Figure 2. One-Class Classification Approach

target class, only data samples of the target microphone are required during training process of OCC model. Each microphone requires a model to be trained on its own data samples, therefore, to identify n microphones, n OCC models are constructed. An example of OCC model construction for a microphone is demonstrated in Fig. 2.

Fig. 2(a) shows a tight decision boundary surrounding *target* microphone data using OCC approach, which can separate it from *outlier* data samples. Here, the data samples of *outlier* classes are shown for demonstration purpose, not input to OCC model. When new microphones become available as shown in Fig. 2(b), the existing OCC model is not required to be re-trained. As a result, the updating process of microphone authentication system is simpler and cheaper than supervised learning approach. Despite the fact that the OCC approach has been successfully employed in several audio forensic tasks such as *sound classification* [6], [7], *scene classification* [8], and *speaker verification* [9], so far no work has been found in the field of microphone forensics where its application scenario is considered appealing.

The objective of this paper is to introduce the application of OCC into the field of Microphone Forensics. Our goals are to evaluate OCC algorithms performance whether they can accomplish the task of identifying and verifying microphone models, and which OCC algorithm can achieve the best results. Hereafter, we use the term “*microphone identification*” to refer to both *identification* and *verification* tasks of microphone forensic category. In order to achieve such goals, we apply a range of relatively new OCC algorithms to audio samples collected from different scenarios using a set of digital microphones. Empirical results and analysis are promising to provide forensic practitioners with an overview about OCC approach. This can support researchers in this area to make microphone forensics a more practical science.

III. DATA COLLECTION AND PRELIMINARY ANALYSIS

A. Recording Devices

In this study, we collected a set of digital recording devices with built-in microphone to collect audio samples. Since, the aim of this work is microphone models detection, thus, the collected devices should be from different models. Otherwise, the audio samples recorded from identical devices are only considered as from a

single microphone model. Besides, a target microphone model is usually identified among a number of available models in *microphone identification* task [1], [3], hence, the experiment in this study should be performed with more than 2 microphone models. Totally, 5 devices are gathered as shown in Table I.

TABLE I.
A SET OF COLLECTED MICROPHONES

Microphone (M_i)	Device Type	Model	Manufacturer
M_1	Audio Recorder	NWZ-B142F	Sony
M_2	Audio Recorder	LX100	Creative
M_3	Audio Recorder	GoGear Mix	Philips
M_4	Camera	PowerShot G5	Canon
M_5	Camera	OptioS6	Pentax

These microphones were used to collect audio samples at different locations such as *indoor*, *quiet park* and *busy street* between 9am to 5pm to form an audio corpus for our study. All audio samples were recorded as a mono signal at 8kHz sampling frequency with 16-Bit quantization. Note that, it would be interesting to perform experiments with more microphone models, however, the use of 5 devices should be enough for model evaluation purpose as it satisfies the natural situation of *microphone identification* task.

B. Audio Features

Audio features are mathematical representations reflecting characteristics of audio signal that are used in statistical pattern recognition based approach for audio forensics. Depending on the tasks of audio forensic practitioners, different sets of audio features could be employed. In [10], AlQahtani et al. made use of MPEG-7 audio low level descriptors along with temporal zero crossing as features vector for automatic recognition of environment sounds. Recently, Sen et al. proposed a new feature extraction technique coming from a new transformation which is based on the *Nyquist* filter bank and achieved significant result in speaker identification [11]. Besides, feature sets extracted from *Linear Predictive coefficients* (LPC) and *mel-frequency cepstral coefficient* (MFCC) also have powerful descriptive capability which are used frequently in gunshot detection [12], audio clips classification [13], environment sound recognition [14], and emotion recognition from speech [15].

In microphone forensic area, different feature sets have been tested in identifying microphone models of recorded audio samples, which include features in time domain, frequency domain and Mel-cepstrum domain. Recent studies conducted by Garcia-Romero and Espy-Wilson [4], and Kraetzer et al. [5] have confirmed that MFCCs are among the best candidates for microphone identification due to its ability of capturing microphone characterizes as well as low dimensionality of 13 coefficients (features). Therefore, we employ MFCCs as our audio features to assess the performance of OCC algorithms in this study. A detail description of MFCCs can be obtained from [16]. From recording signal of the

microphones, MFCCs are extracted from randomly sampled audio segments of 0.25 seconds in length, to form data instances with 13 vector features. Data instances extracted from a microphone M_i are assigned with the label of that microphone.

C. Preliminary Analysis

In this section, we would like to evaluate the effectiveness of the selected audio features (MFCC), to see if it can really describe the difference between audio signal recorded by different microphones. A standard technique to perform this task is to use Multi-Dimensional Scaling (MDS) [17], a multivariate and exploratory data analysis technique. The input data of MDS is a measure of global similarity or dissimilarity of objects (data instances) under investigation, and the outcome is a spatial configuration, where the objects are repressed as points. Similar objects are represented by point closed to each other, dissimilar objects by points that are far apart. Using this technique, we can visualize our multidimensional data set to see the distribution of data that was recorded from the microphones. Fig. 3 shows the results of MDS on data instances of 5 microphones (500 instances each) in case of *low noise* and *high noise* signal, corresponding to *indoor* and *busy street* environment.

Fig. 3(a) for *indoor* recording shows that instances belong to the same microphone tend to stay closed to each other and form a cluster. The clusters of different microphones are relatively easy to identify visually as they are quite-well separated from each others. This proof that the MFCC features are good candidate for presenting audio signal for microphone identification task. Fig. 3(b) for *busy street* recording shows that, even though, data instances of the same microphone are still formulating a cluster, there is some overlap between clusters of microphones. This may be due to high noise level of *busy street* environment. Due to the nature of recording signal, microphone identification task can be easier to perform for low noise recording, it will be harder to do so for recording in noisy environment.

Considering the fact that most of the existing approaches generate data samples by randomly selecting audio segments of t seconds in length (0.25 second in our case), then features are extracted as input to classifier. When data samples are extracted from noisy recordings, many high noise level audio segments are taken into consideration, which significantly affect classification performance. Therefore, a better data sampling technique and classification approach for microphone identification problem are required so that the impact of noise to classification algorithm can be reduced. The method to tackle this problem will be addressed clearly in section IV-B.

IV. METHODOLOGY

In this section, we first outline several OCC algorithms which are used in this study, then, we present our representative instances identification framework (RICF) to improve OCC performance for noisy recording signal.

A. One-class Classification algorithms

The area of OCC is considerably well adapted to the problems of microphone identification where sampling every microphone model is an impossible task. For OCC approach, a boundary around the well-sampled target distribution is constructed that may discard a small percentage of target examples, and consequently have the potential to be able to identify majority samples of target microphone while throwing out as many examples of other microphones as possible. A sample is classified as member of target class if its assigned score by OCC model lies above a given threshold, otherwise, it is considered belonging to an outlier class.

In the context of this study, we use a number of OCC algorithms implemented in *Data Description toolbox* by Tax [18] which are described as follows:

- **One-class Gaussian Model (1-GN):** uses a simple Gaussian to characterize target class. When instances of target class are input to this algorithm, a model is constructed using one Gaussian distribution, where μ is estimated mean and Σ is estimated covariance matrix of target “points”:

$$f(x) = (x - \mu)^T \Sigma^{-1} (x - \mu) \quad (1)$$

An instance x is classified as target class if $f(x)$ is greater than a specified target error θ , otherwise, it is considered belonging to outlier class. The threshold value θ can be supplied by the user, otherwise, it is assigned with a default value in the OCC toolbox.

- **One-class Gaussian Mixture Model (1-GNM):** uses a mixture of K Gaussians to construct a more flexible description for target class. The model can be presented as follows:

$$f(x) = \sum_{i=1}^K P_i \exp(-(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)) \quad (2)$$

In this model, EM algorithm is utilized to optimize the parameters P_i , μ_i and Σ_i .

- **One-class K-Means (1-KM):** describes the target data by k clusters with center c_i to be estimated such that the average distance to a cluster center is minimized. The model for 1-KM is described as follows:

$$f(x) = \min_i (x - c_i)^2 \quad (3)$$

- **One-class K-Nearest Neighbors (1-KNN):** evaluates a new object x by computing the distance to its k nearest neighbors in target class. Then this distance is compared to a threshold to evaluate the prediction output of x .
- **One-class Principal Component Analysis (1-PCA):** describes the target data by a linear subspace W which is defined by the k eigenvector of covariance matrix of target class data. The fitness of new object x to the target subspace is evaluated by comparing it to reconstruction error:

$$f(x) = \|x - x_{proj}\|^2 \quad (4)$$

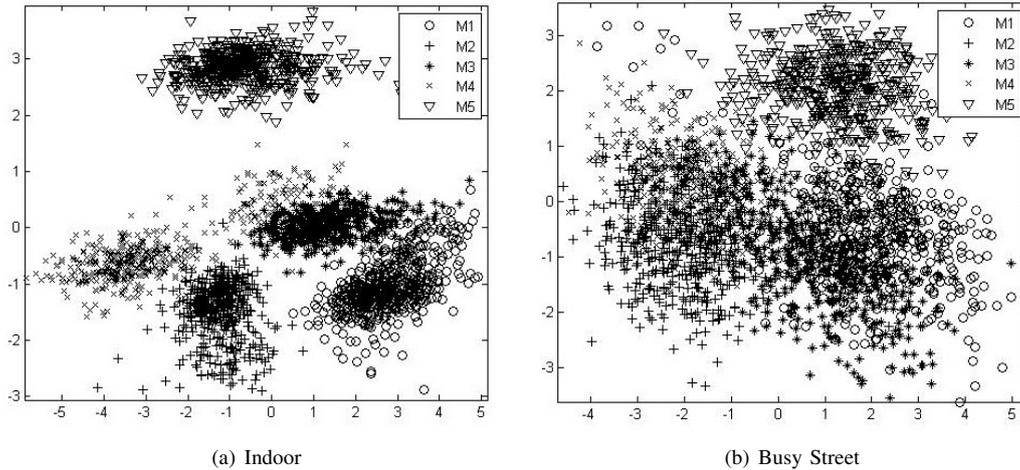


Figure 3. MDS result on sample data instances of 5 microphones

where the projection x_{proj} is calculated by:

$$x_{proj} = W(W^T W)^{-1} W^T x \quad (5)$$

- **One-class Incremental Support Vector Machine (1-ISVM):** fits a hypersphere around the target class without using an external quadratic programming optimizer or kernel. The employment of Support Vector Machine technique into context of one-class was originally proposed by Tax and Duin [19], which is known as *Support Vector Domain Description (SVDD)*. It was proven to achieve significant results in application to audio surveillance system [20] and speaker verification [9]. Therefore, we would like to evaluate this technique in this study using 1-ISVM as a more advanced version of SVDD.

B. Representative Instances Classification Framework

In this section, we present RICF method which is designed to reduce the impact of noise, and can thus improve OCC performance for noisy recordings. It contains two steps: 1) Representative Training Instances Extraction; 2) Bag of Instances Classification.

1) *Representative Training Instances Extraction:* This section present an representative instances identification method for sampling training data set, so that the impact of noise is reduced. From preliminary analysis in section III-C, we notice that instances of the same microphone tend to stay closer to each others and form a cluster, for both *low noise* and *high noise* cases. An audio segment of microphone M_i is considered to have less noise, if the data instance extracted from it stays closer to cluster centroid C_{M_i} of the cluster which is formulated by data samples of M_i . Given an audio record R_{M_i} of microphone M_i (T seconds in length), and r_1, r_2, \dots, r_m are segments of t seconds extracted from R_{M_i} ($t < T$). the best representative instance is the instance that is extracted from a segment $r_{RI} \in R_{M_i}$ which is the closest to a cluster center C_{M_i} .

Since, the cluster of each microphone data samples can be recognized visually based on its density of distribution

as shown in section III-C, a method comes naturally to identify cluster center of each microphone data is to use *Density Estimation (DE)* [21]. An advantage of DE is that the estimation of cluster center is subjected to probability distribution, and thus robust to noise. In this study, we use *Kernel Density Estimation (KDE)* [22] with *Gaussian Kernel*, a non-parametric DE technique, to find densest region of each microphone data samples and consider it as cluster center C_{M_i} . Here, the data samples for estimating C_{M_i} are the training data instances that are randomly samples from training set. They are not considered as training instances, but instead used to find the center representing the cluster of microphone data distribution.

To identify representative instance, a sliding window of t seconds is moved along each audio record R_{M_i} for m steps, thus, m instances I_1, I_2, \dots, I_m are extracted from m segments r_1, r_2, \dots, r_m of sliding window. Representative instance I_{RI} for each audio record R_{M_i} is the instance that satisfy $RI = \arg \min_n dis(C_{M_i}, I_n)$, where $dis(C_{M_i}, I_n)$ is the *euclidian distance* between I_n and cluster center C_{M_i} . This process is repeated for each audio record of each microphone to construct a representative training data set for each microphone model.

In this study, the length of each audio record is 2 seconds, and sliding window is 0.25 second. An example of sliding windows is shown in Fig. IV-B.1.

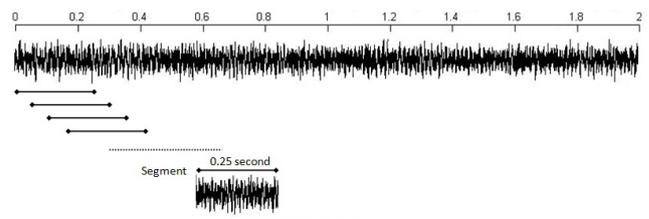


Figure 4. An example of sliding window for an audio record

The sliding window moves with a step of 0.0625 second. Thus, 36 segments are extracted from each 2 seconds record, which will be further represented by 36

instances with 13 vector features. Then, only 1 instance is selected as representative instance to represent each audio record in training process.

2) *Bag of Instances Classification*: After representative instances are extracted for each microphone, the training process is then performed with OCC algorithm, a OCC model is built for each microphone to characterize its target training samples. Note that, the process of identifying representative instances from training audio records is based on prior knowledge of which microphone it belongs to. However, no prior knowledge is given about an audio record in testing stage. It is impossible to extract a particular representative instance from a set of instance extracted from such testing record. To classify a new audio record, all instances extracted from each audio record are stored in a bag of instance $B = \{I_1, I_2, \dots, I_m\}$ and input to trained OCC models. Each trained OCC model o_i of each microphone M_i generates a set of predicted labels $L_i = \{l_{i1}, l_{i2}, \dots, l_{im}\}$ for each bag B . The predicted label of bag B is assigned to the microphone label whose model predicts the most number of instances in B as its target class. In RICF, the testing unit is an audio record containing m segments, and the input data unit to OCC models is a bag of instances, rather than individual segment/instance as in traditional OCC framework.

V. EXPERIMENT AND ANALYSIS

This section first describes our evaluation metric to assess the performance of OCC algorithms. Then, detailed analysis of experiment results is presented, together with discussion and implication analysis.

A. Evaluation Metrics

At the training stage, one model o_i of OCC algorithm O is built for each microphone, thus, m microphones require m OCC models o_1, \dots, o_m to be built. Then, they are applied to the test set to detect audio samples belonging to their target class. Since we are only interested in the absolute performance of OCC algorithms on “target” data, *recall* rate and *precision* rate are used to measure their detection capability. Here, *Recall* refers to the fraction of the target class data instances classified correctly as target class, *precision* refers to the fraction of the outlier data instances classified incorrectly as target class.

Suppose there are P audio samples of target class and N audio samples of outlier class in the test set for microphone M_i , among which P_a target samples are recognized correctly, and N_r outliers are recognized incorrectly as target class, then the *recall* and *precision* of OCC model o_i on M_i are calculated as:

$$recall = \frac{P_a}{P} \tag{6a}$$

$$precision = \frac{N_r}{N} \tag{6b}$$

To evaluate the overall performance of O , we use *overall recall* and *overall precision* which are averages

of *recall* values and *precision* values over m models o_1, \dots, o_m of O . In a classification task, the prediction result is expected to be better than a random guess (above 0.5 for *recall* score). The *recall* is considered to be higher if it is closer to 1.0.

In the experiment, the OCC algorithms mentioned in section IV will be used. Parameters of each classifier are kept as default values in the OCC toolbox [18]. On the training set, the rejection rate of the target class was set to 10% in order to provide a tight decision boundary around the target class.

B. Result Analysis

To assess the performance of OCC for microphone detection, we perform three major experiments with audio samples at different noise levels corresponding to different recording locations. In these cases, the training sets are sampled to be equal to testing set of 1000 instances for each microphone. Considering the fact that noisy environment is usually more difficult than quiet environment, we perform another experiment for the case of *busy street* with different number of training instances to examine if the performances of OCC algorithms can be improved by incorporate more training samples.

Then, RICF method will be evaluated against the traditional method in the case of *busy street*. The difference of OCC when using with RICF is that training set is representative instance extracted from audio records, and test set is bags of instances.

1) *Microphone Identification for indoor recordings*: The recorded audio samples in this case were taken from *indoor* environment such as small room, big room, building and lecture theatre where little background noise is presenting.

All 6 OCC algorithms were trained and tested on the same test set extracted previously. The experiment results are shown in Table II.

TABLE II. MICROPHONE IDENTIFICATION RESULT FOR INDOOR ENVIRONMENT RECORDINGS

Algorithm		M_1	M_2	M_3	M_4	M_5	Overall
1-GN	Recall	0.873	0.892	0.897	0.685	0.867	0.843
	Precision	0.000	0.000	0.000	0.000	0.000	0.000
1-GNM	Recall	0.830	0.800	0.858	0.543	0.842	0.774
	Precision	0.000	0.000	0.000	0.000	0.000	0.000
1-KM	Recall	0.885	0.881	0.896	0.623	0.893	0.836
	Precision	0.000	0.000	0.000	0.000	0.000	0.000
1-KNN	Recall	0.867	0.894	0.896	0.626	0.897	0.836
	Precision	0.000	0.000	0.000	0.000	0.000	0.000
1-PCA	Recall	0.908	0.807	0.897	<u>0.767</u>	0.915	0.859
	Precision	0.061	0.052	0.003	0.148	0.000	0.053
1-ISVM	Recall	0.907	0.818	0.916	<u>0.750</u>	0.872	0.853
	Precision	0.000	0.000	0.000	0.000	0.000	0.000

In general, all algorithms achieved high *overall recall* in identifying target microphones as indicated by the values of over 0.8, only except for 1-GNM with lower detection rate of 0.744. Importantly, almost no error was made

across the models as indicated by *overall precision* value of 0, only a small error is shown for 1-PCA.

As each microphone M_i is identified by a unique model o_i , the assessment of OCC algorithm O needs to be considered together with the performance of each individual model corresponding to each microphone. All six algorithm achieved high *recall* and no *precision* in detecting microphones M_1 , M_2 , M_3 and M_5 , while it is harder to recognize microphone M_4 with significantly lower *recall* values. Even so, 1-PCA and 1-ISVM achieve considerably good *recall* for M_4 of above 0.750, especially for 1-ISVM which has no error at all.

These results indicate that the OCC algorithms are suitable for detecting and verifying microphone models in indoor environment. Further evaluation for OCC will be carried out in the next experiment with more noisy audio samples.

2) *Microphone Identification for quiet park recordings*: In this case, the audio records were taken from a *quiet park* with considerable background noise. Training and testing data sets were both extracted from these audio records, and then similar evaluation process was applied with the same parameter values for each algorithm. The results are shown in Tables III.

TABLE III.
MICROPHONE IDENTIFICATION RESULT FOR QUIET PARK
ENVIRONMENT RECORDINGS

Algorithm		M_1	M_2	M_3	M_4	M_5	<i>Overall</i>
1-GN	Recall	0.750	0.916	0.956	0.911	0.893	0.885
	Precision	0.001	0.090	0.001	0.174	0.008	0.055
1-GNM	Recall	<u>0.506</u>	0.751	0.822	0.756	0.838	0.735
	Precision	0.000	0.016	0.000	0.029	0.000	0.009
1-KM	Recall	0.768	0.907	0.926	0.897	0.923	0.884
	Precision	0.001	0.056	0.003	0.135	0.001	0.039
1-KNN	Recall	0.648	0.810	0.866	0.820	0.887	0.806
	Precision	0.000	0.024	0.000	0.032	0.000	0.011
1-PCA	Recall	0.795	0.831	0.932	0.873	0.841	0.854
	Precision	0.229	0.121	0.002	0.201	0.069	0.124
1-ISVM	Recall	0.871	0.969	0.954	0.973	0.878	0.929
	Precision	0.003	0.300	0.221	0.353	0.066	0.188

It is interesting to see that the *overall recall* of the algorithms are not significantly different from previous experiment. However, we notice that more errors were made as indicated by higher *overall precision* values. In particular, 1-SVM outperformed other algorithm in detecting microphones with *overall recall* of 0.929, however, it also made signification number of wrong prediction as indicated by *overall precision* value of 0.188.

In view of individual detection model, none of them failed to detect their target microphones as shown by high *recall* and low *precision* values for $M_1 - M_5$. Even for 1-GNM scoring lowest *recall* of 0.506 on M_1 , it is still accepted as we are interested in the detection capability for target class of OCC models.

Despite some errors were made, OCC algorithms were able to identify microphone models for audio samples taken from outdoor (*quiet park*) environment. Further evaluation of OCC approach will be performed for extremely noisy audio records in the next section.

3) *Microphone Identification for busy street recordings*: In this experiment, the audio samples were obtained from a *busy street* environment with significant amount of noise presenting. Audio features were extracted and input to OCC algorithms as in previous experiments. The detection results are shown in Table IV.

TABLE IV.
MICROPHONE IDENTIFICATION RESULT FOR BUSY STREET
ENVIRONMENT RECORDINGS

Algorithm		M_1	M_2	M_3	M_4	M_5	<i>Overall</i>
1-GN	Recall	0.690	<u>0.311</u>	<u>0.291</u>	0.679	0.681	0.530
	Precision	0.040	0.006	0.005	0.152	0.084	0.057
1-GNM	Recall	0.320	0.143	0.188	0.306	0.331	<u>0.257</u>
	Precision	0.011	0.001	0.002	0.047	0.011	0.014
1-KM	Recall	0.555	<u>0.481</u>	<u>0.384</u>	0.514	0.625	0.512
	Precision	0.108	0.016	0.036	0.111	0.092	0.073
1-KNN	Recall	0.175	0.149	0.185	0.259	0.262	<u>0.206</u>
	Precision	0.007	0.001	0.003	0.026	0.006	0.008
1-PCA	Recall	0.787	<u>0.263</u>	<u>0.332</u>	0.743	0.688	0.563
	Precision	0.247	0.082	0.132	0.435	0.234	0.226
1-ISVM	Recall	0.804	0.868	0.506	0.876	0.865	0.784
	Precision	0.230	0.027	0.081	0.322	0.255	0.183

From Table IV, we can see that 1-GNM and 1-KNN failed in identifying their target microphones as indicated by *overall recall* of below 0.5. Although, the *overall recall* values of 1-GN, 1-KM and 1-PCA are higher than 0.5, they failed to detect microphone M_2 and M_3 as shown by *recall* values of less than 0.5.

On the contrary, 1-ISVM outperformed all others and was able to detect all microphone models with *overall recall* of 0.784, especially, none of its individual models has *recall* value lower than 0.5. Even though, it also make significant wrong prediction as shown by *precision* of 0.183.

4) *Effects of training sample size*: As the matter of fact that, most OCC algorithms was failed to identify their target microphone for noisy environment as presented in previous section, this is due to the present of noise in the evaluated data set. A traditional method to lower the impact of noise is to increase the number of training instances. In this experiment, we assess the performance of OCC algorithms in case of noisy environment (*busy street*) with increasing number of training samples in each iteration, to examine if bigger training set can improve their performance. The training set was started from 1000 instances and increased by 100 instances in each iteration. In this case, we would like to evaluate the overall performance of those algorithms, therefore, only *overall recall* and *overall precision* values are presented, as shown in Fig. 5.

From Fig. 5(a), we can see that, generally, the *overall recall* of OCC algorithms are increasing steadily as the number of training instances increasing. Whilst, 1-GN, 1-KM and 1-PCA are stabilizing from training size of 2400 instances, the *overall recall* values of 1-GNM and 1-KNN continue to growth slightly. In contrast, no significant improvement is found for 1-ISVM as its detection rate stays around 0.8 even when the training set is tripled to 3000 instances.

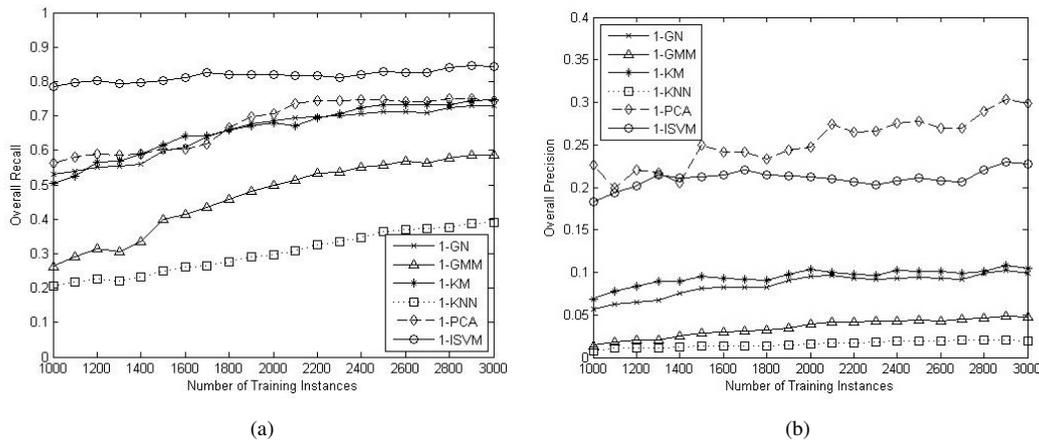


Figure 5. Effects of different training sample sizes

In Fig. 5(b), rough growth in *overall precision* rate is found for 1-PCA, while that of 1-SVM is increasing gradually until training size reach 1300 instances then stylizing between 0.2 and 0.25. Similarly, the *precision* rates of 1-GA and 1-KM increase slightly until training size reach 2000 instances then become stable of around 0.1. At the same time, precision rates of 1-GNM and 1-KNN are also increased but remained lower than 0.05 even when the training data is tripled.

In summary, the above result shows that the increase of training size can slightly improve performance on most OCC algorithms on noisy audio recordings. However, their performance is still far from acceptable standard due to low *recall* and high *precision* rate, and most OCC algorithms are still failed to identify target microphones from noisy recordings. An alternative method should be developed that can help OCC algorithms to overcome the problem of noisy data to make it more suitable for real life applications. This issue will be addressed by RICF method in the following section with experiment result and comparison analysis.

5) *RICF evaluation on Microphone Identification for busy street recordings*: For each microphone in this experiment, the training data set was made up by 1000 representative instances extracted from 1000 audio recording of 2 second in length, following the method presented in section IV-B.1. The testing data set were 1000 bags of instances extracted from testing audio records. OCC algorithms were trained on representative instances then applied to testing bags as described in section IV-B.2.

To make it easy for comparison, we also include results of OCC algorithms for both with and without RICF as shown in Table V.

From Table V, we can see that there is significant improvement for OCC algorithms (1-GN, 1-GNM, 1-KM, 1-KNN) with RICF in identifying target microphone of audio records as indicated by *overall recall* of above 0.8. With high *recall* rate and low *precision* for both overall and individual rate, there four OCC algorithms with RICF are considered successful in performing microphone identification for noisy environment.

TABLE V.
MICROPHONE IDENTIFICATION RESULT FOR BUSY STREET ENVIRONMENT RECORDINGS WITH AND WITHOUT RICF

Alg.	RICF		M_1	M_2	M_3	M_4	M_5	Overall
1-GN	No	Rec.	0.690	0.311	0.291	0.679	0.681	0.530
		Pre.	0.040	0.006	0.005	0.152	0.084	0.057
	Yes	Rec.	<u>0.987</u>	<u>0.807</u>	<u>0.887</u>	<u>0.861</u>	<u>0.828</u>	0.874
		Pre.	0.107	0.005	0.012	0.028	0.006	0.032
1-GNM	No	Rec.	0.320	0.143	0.188	0.306	0.331	0.257
		Pre.	0.011	0.001	0.002	0.047	0.011	0.014
	Yes	Rec.	<u>0.988</u>	<u>0.833</u>	<u>0.871</u>	<u>0.818</u>	<u>0.806</u>	0.863
		Pre.	0.134	0.006	0.009	0.016	0.006	0.034
1-KM	No	Rec.	0.555	0.481	0.384	0.514	0.625	0.512
		Pre.	0.108	0.016	0.036	0.111	0.092	0.073
	Yes	Rec.	<u>0.946</u>	<u>0.824</u>	<u>0.826</u>	<u>0.800</u>	<u>0.786</u>	0.836
		Pre.	0.130	0.013	0.023	0.023	0.017	0.041
1-KNN	No	Rec.	0.175	0.149	0.185	0.259	0.262	0.206
		Pre.	0.007	0.001	0.003	0.026	0.006	0.008
	Yes	Rec.	<u>0.952</u>	<u>0.858</u>	<u>0.867</u>	<u>0.848</u>	<u>0.829</u>	0.871
		Pre.	0.101	0.008	0.021	0.022	0.010	0.032
1-PCA	No	Rec.	0.787	0.263	0.332	0.743	0.688	0.563
		Pre.	0.247	0.082	0.132	0.435	0.234	0.226
	Yes	Rec.	0.714	0.340	0.371	0.494	0.731	0.530
		Pre.	0.203	0.058	0.119	0.157	0.050	0.117
1-ISVM	No	Rec.	0.804	0.868	0.506	0.876	0.865	0.784
		Pre.	0.230	0.027	0.081	0.322	0.255	0.183
	Yes	Rec.	0.937	0.814	0.726	0.707	0.523	0.741
		Pre.	0.188	0.028	0.069	0.028	0.011	0.065

Besides, no significant difference was found for 1-PCA and 1-ISVM as the *overall recall* of OCC with RICF is slightly lower than without RICF but it also made less mistake as shown by lower *overall precision* rate. Although, no conclusion can be drawn at this stage if RICF can improve the performance of 1-PCA and 1-ISVM, it is clear that 5 out of 6 tested OCC algorithms with RICF have been able to perform microphone identification for noisy signal (except for 1-PCA). This result is better than the use of OCC without RICF, when 5 out of 6 failed to perform this task as in section V-B.3.

C. Discussion

The experiment results in section V-B.1 and V-B.2 support the claim that *One Class Classifier* (OCC) is a

suitable approach to microphone forensics. Most algorithms achieve high performance on audio samples that are recorded from little noise environment, as the recorded signals are not influenced significantly by outside signal, and can thus reflect better microphone characteristics. The quality of OCC algorithms are reduced when applying to noisy audio signal as presented in V-B.3, most OCC algorithms were failed to identify target microphone of audio recordings, even if more training samples are provided as shown in V-B.4. However, the problem of noisy signal can be reduced effectively by apply OCC algorithms together with RICF framework, experiment results in section V-B.5 have proven its advantage with the succeed of 5 among 6 tested algorithms. Since, 1-PCA failed in above experiments of busy street scenario, its may not be a suitable candidate for microphone identification with high noise level signal.

The key advantage of the OCC approach is that they focus on the detection of target class instances by using tight decision boundary obtained from training data. As consequence, only a small number of fault detection cases are made as indicated by low *precision* rate across the OCC models in previous experiments. In addition, these OCC models only need to be trained once, and are independence from each other. In other words, when a new microphone become available, only a new OCC model trained on audio samples of that microphone is required without retraining existing OCC models. For such reasons, the approach using OCC is more practical, cost effective and easier for system maintenance than traditional supervised learning.

On the other hand, a drawback of OCC approach is that it requires a classification model to be built for each microphone. Therefore, more space is required to store the trained OCC models than the *supervised learning* approach, which only needs a unique classification model for all available microphones, Even though, the sizes of OCC classification models are usually small, thus, the required space would not be an issue to computer storage capacity nowadays.

VI. CONCLUSIONS

Microphone identification and verification are important tasks in integrity and authenticity assurance of information which are becoming more and more crucial nowadays in criminal investigation. However, only a limited number of approaches have been used in the literature which utilize the machine learning methods to support for microphone forensic practitioners in doing this work. Current applications of supervised learning are still facing huge challenges due to time and associated cost of collecting audio samples from a large number of microphones for model training purposes; In addition, frequent retraining of existing classification models is required. In this paper, we present the first attempt in automated microphone detection using One-Class Classification approach which exhibits to be effective to alleviate the challenges of microphone forensics.

To be precise, we presented an evaluation of 6 relatively new OCC algorithms in detecting microphone models under different conditions of noise level. Experiment results indicate that the tested OCC algorithms are able to detect microphone model with high recall and low precision rate. Moreover, We have also developed a representative classification framework RICF which can effectively support for OCC algorithms to improve their performance in tackling the problem of noisy signal. In summary, the contributions of this study are the introduction of OCC approach into microphone forensic which address the limitation of prior works, and the proposing of RICF method to tackle noisy signal problems. With the proposed techniques, the task of microphone identification have been made possible for application in real life situation.

A natural extension of this work will be testing OCC algorithms on a wide range of microphone models and under more sophisticated recording scenarios. Furthermore, we will investigate the extent to which a real time microphone forensic system could be developed for online usage that can support audio forensic practitioners in performing their tasks.

ACKNOWLEDGMENT

The authors would like to thank Professor Lynn Batten, of School of Information Technology at Deakin University, for proposing and funding this research project.

REFERENCES

- [1] C. Kraetzer, A. Oermann, J. Dittmann, and A. Lang, "Digital Audio Forensics : A First Practical Evaluation on Microphone and Environment Classification," in *Proceedings of the 9th workshop on Multimedia and security*, Dallas, Texas, September 2007, pp. 63–74.
- [2] C. Kraetzer, M. Schott, and J. Dittmann, "Unweighted fusion in microphone forensics using a decision tree and linear logistic regression models," in *Proceedings of the 11th workshop on Multimedia and security*. Princeton, New Jersey, USA: ACM Press, September 2009, pp. 49–56.
- [3] R. Buchholz, C. Kraetzer, and J. Dittmann, "Microphone Classification Using Fourier Coefficients," *Lecture Notes in Computer Science*, vol. 5806, pp. 235–246, 2009.
- [4] D. Garcia Romero and C. Y. Espy Wilson, "Automatic acquisition device identification from speech recordings," in *Proceeding of IEEE International Conference on Acoustics Speech and Signal Processing*, Dallas, Texas, US, March 2010, pp. 1806–1809.
- [5] C. Kraetzer, K. Qian, M. Schott, and J. Dittmann, "A context model for microphone forensics and its application in evaluations," *Media Watermarking, Security, and Forensics III*, vol. 7880, p. In Press, 2011.
- [6] A. Rabaoui, H. Kadri, Z. Lachiri, and N. Ellouze, "One-Class SVMs Challenges in Audio Detection and Classification Applications," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 1–14, 2008.
- [7] A. Rabaoui, M. Davy, S. Rossignol, and Z. Lachiri, "Improved one-class svm classifier for sounds classification," in *Proceeding of IEEE Conference on Advanced Video and Signal Based Surveillance*, London, United Kingdom, September 2007, pp. 117–122.

[8] G. FengJuan, S. ShuQian, and W. XiaoHui, "Using One-Class SVMs and MP for Audio Recognition of Action Scenes," in *Proceeding of 2nd International Workshop on Education Technology and Computer Science*, Wuhan, China, March 2010, pp. 401–404.

[9] A. Brew, M. Grimaldi, and P. Cunningham, "An evaluation of one-class classification techniques for speaker verification," *Artificial Intelligence Review*, vol. 27(4), pp. 295–307, 2008.

[10] M. O. AlQahtani, G. Muhammad, and Y. A. Alotaibi, "Environment Sound Recognition using Zero Crossing Features and MPEG-7," in *Proceeding of Fifth International Conference on Digital Information Management*, Thunder Bay, Ontario, Canada, July 2010, pp. 502–506.

[11] N. Sen, T. K. Basu, and H. A. Patil, "Significant Improvement in the Closed Set Text-Independent Speaker Identification Using Features Extracted from Nyquist Filter Bank," in *Proceeding of 5th International Conference on Industrial and Information Systems*, India, Jul 2010, pp. 303–308.

[12] I. L. Freire and J. A. Apolinario Jr., "Gunshot detection in noisy environments," in *Proceeding of the 7th International Telecommunications Symposium*, Manaus, Brazil, September 2010.

[13] P. Dhanalakshmi, S. Palanivel, and V. Ramalingam, "Classification of audio signals using AANN and GMM," *Applied Soft Computing*, vol. 11, no. 1, pp. 716–723, Jan. 2011.

[14] M. O. Alqahtani and A. S. Al mazyad, "Environment Sound Recognition for Digital Audio Forensics Using Linear Predictive Coding Features," in *Proceeding of International Conference on Digital Information Processing and Communications*, Ostrava, Czech Republic, July 2011, pp. 301–309.

[15] J. Rong, G. Li, and Y.-P. P. Chen, "Acoustic feature selection for automatic emotion recognition from speech," *Information Processing and Management*, vol. 45(3), pp. 315–328, 2008.

[16] J. Ye, "Speech recognition using time domain features from phase space reconstructions," Ph.D. dissertation, Marquette University, Milwaukee, Wisconsin, May 2004.

[17] W. Florian, "An introduction to mds," Aalborg University, Denmark, Tech. Rep., 2003.

[18] D. Tax, "Data description toolbox dd tools 1.7.5," Delft University of Technology, Delft, The Netherlands, Tech. Rep., May 2010.

[19] D. M. Tax and R. P. Duin, "Support vector domain description," *Pattern Recognition Letters*, vol. 20, pp. 1191–1199, 1999.

[20] A. Rabaoui, M. Davy, S. Rossignol, and N. Ellouze, "Using one-class svms and wavelets for audio surveillance," *IEEE Transactions on Information Forensics and Security*, vol. 3(4), pp. 763–775, 2008.

[21] T. Hastie, R. Tibshirani, and J. Friedman, Eds., *The Elements of Statistical Learning*. New York: Springer, 2001, ch. 6.

[22] G. R. Terrell and D. W. Scott, "Variable kernel density estimation," *The Annals of Statistics*, vol. 20(3), 1992.



Shaowu Liu is an honours student at Deakin University. He received his Bachelor of Computer Science degree from Deakin University in 2011. His research interests include machine learning, applications of data mining techniques to problems in tourism, social media and scientific fields.



Xinghua Yang received MEng degree from the Graduate School of Chinese Academy of Sciences (CAS) in 2011. He is now working in the High Performance Network Laboratory, in the Institute of Acoustics, CAS as a research intern. His research interests include mobile internet, cognitive network, pattern recognition and machine learning.



Zhi Li holds her PhD in IT. Her PhD research areas were active queue management and quality of service. Her PhD dissertation is entitled "Fuzzy Logic Based Robust Control of Queue Management and Optimal Treatment of Traffic over TCP/IP Networks". She worked as a research assistant, tutor and marker. She has accumulated strong computer skills and is proficient in programming with C/C++, MATLAB, C#. Currently, she joins Logicalis Australia and work as a network engineer. To date, she has published ten research papers and one book chapter.



Yongli Ren received his B.S in 2006, and his Master degree in 2009, from Zhengzhou University, China. He is currently a PhD candidate at School of Information Technology, Deakin University. His current research interests include recommender system, pattern recognition, and data mining



Huy Quan Vu received his Bachelor of Information Technology (honours) and completed an internship at CSIRO - Mathematics, Informatics and Statistics in 2010. He is currently a PhD candidate at Deakin University. His research interests include time series data mining, sentiment mining, text mining, and data mining application in tourism.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.