

An Automatic Pipeline Monitoring System Based on PCA and SVM

C. Wan, and A. Mita

Abstract—This paper proposes a novel system for monitoring the health of underground pipelines. Some of these pipelines transport dangerous contents and any damage incurred might have catastrophic consequences. However, most of these damage are unintentional and usually a result of surrounding construction activities. In order to prevent these potential damages, monitoring systems are indispensable. This paper focuses on acoustically recognizing road cutters since they prelude most construction activities in modern cities. Acoustic recognition can be easily achieved by installing a distributed computing sensor network along the pipelines and using smart sensors to “listen” for potential threat; if there is a real threat, raise some form of alarm. For efficient pipeline monitoring, a novel monitoring approach is proposed. Principal Component Analysis (PCA) was studied and applied. Eigenvalues were regarded as the special signature that could characterize a sound sample, and were thus used for the feature vector for sound recognition. The denoising ability of PCA could make it robust to noise interference. One class SVM was used for classifier. On-site experiment results show that the proposed PCA and SVM based acoustic recognition system will be very effective with a low tendency for raising false alarms.

Keywords—One class SVM, pipeline monitoring system, principal component analysis, sound recognition, third party damage.

I. INTRODUCTION

MANY important pipelines in modern cities are laid underground and are often referred to as lifeline infrastructures. Recent reports in literature [1] show that most damages to these pipelines are usually caused by third party activities, rather than material failure and corrosion; for example, surrounding construction activities. Although unintentional damages might be considered rare events, a single incident may have catastrophic consequences; the Ghislenghien gas pipeline explosion disaster in Belgium on July 30, 2004 resulted in 24 deaths and over 120 injuries.

For pipelines that are deemed dangerous, such as those carrying gas and high pressure oil, preventive measures to detect potential threats are more important than measures to detect real damages. In actual practices, dangerous pipelines are regularly patrolled by special personnel, either on foot, by vehicles, or even by helicopters. However, this kind of manual checking is laborious, economically expensive and is not seen

to be efficient or necessarily effective. Therefore, there is an increasing necessity for automatic, continuous, and low cost pipeline monitoring systems.

In Europe, a monitoring system based on image processing and Unmanned Aerial Vehicles (UAV) technologies is under development [2]. However, this system is expensive and requires the tight integration of several complicated technologies. Furthermore, the tendency for false alarms is still not low enough.

In the United States, Gas Technology Institute (GTI) is developing another monitoring system with the objective of preventing third party damage, particularly that resulting from nearby construction activities [3]. This system uses optical fibers, which are buried between the surface and the pipelines, to detect for vibrations in the ground. The magnitude and profile of the vibrations are then used to determine the existence of construction equipment nearby. However, this system is only applicable in areas where there is nothing on the topsoil. For pipelines under an asphalt or concrete road, which is the most common case in Japan, this system is not effective.

Bearing in mind the availability of these systems, Wan et al. proposed a pipeline monitoring system based on acoustic recognition [4][5]. Construction work near pipelines is identified by the sounds emitted from the construction machines used. By continuously studying the surrounding noise, potential threats to the pipelines can be identified by detecting for dangerous construction equipment in the vicinity of the pipelines. In [4], a 0.2s sound sample was extracted out for recognition from time to time. Mel Frequency Cepstral Coefficient (MFCC) was used as the feature. However, depending on only a 0.2s sample, decisions will not be reliable due to noises. In [5], with the consideration of the noise effect, a tag train based postprocessor was proposed to make the final decision, based on many segmented initially recognized frames. In this paper, however, a different approach will be applied. With PCA and one class SVM, recognition of a whole sound sample, rather than tiny segmented frames, can be achieved. Simultaneously, robustness to noises due to the denoising ability of PCA, and thus accurate decision making, can also be realized. Meanwhile, high performance support vector machines could lead to high recognition accuracies. For these considerations, pipeline monitoring approach based on PCA and one class SVM is studied in detail in this paper.

C. Wan, is with System Design Department, Keio University, 3-14-1, Hiyoshi, Yokohama 223-8522, Japan (phone: +81 45-566-1776; fax: +81 45-566-1776; e-mail: chunfengwan@gmail.com).

A. Mita, is with System Design Department, Keio University, 3-14-1, Hiyoshi, Yokohama 223-8522, Japan (e-mail: mita@sd.keio.ac.jp).

II. PIPELINE MONITORING MECHANISM

Usually, it is the presence of surface construction activities that threaten the well-being of underground pipelines; construction activities along asphalt or concrete roads threaten the pipelines that run along underneath the roads. In such activities, a road cutter is often used before any other equipment. Furthermore, the operation of a road cutter is accompanied with a very loud noise, which makes the recognition feasible and practical. For these two reasons, we can focus on detecting road cutters to determine if there are any construction activities near the pipelines. Fig. 1 shows a road cutter cutting the road.



Fig. 1 A road cutter is cutting the road

A sensor network could be deployed above ground along the pipelines. Each sensor will have a microphone to capture sounds and a small chip to process the sound signals by using acoustic recognition methods. Once it detects a dangerous sound, it will send off an alarm and at the same time send a message to the control center so that immediate and relevant measures can be expediently executed. Obviously, essential to the proposed system is the cutter detection by sound recognition. Nearly all environmental sound recognition researches in literature use short audio clips, often lasting from 1s to 10s. For example, Ma [6] recognized an environment sound based on a 3s sample, while Lu [7] made a Content-based audio classification based on a 1s sample. In our research, a short period sound sample, lasting several seconds, is also applied to make sound recognition. In simplicity, the sound based cutter detection system on the whole can be described in 4 steps:

- 1) Sound capture – catch a sound signal. If the strength of the sound exceeds a threshold, activate the processor to analyze the signal, otherwise do nothing.
- 2) Sample extraction – extract a sound sample from the incoming sound signal.
- 3) Sound classification – process the sample and decide whether belongs to road cutter or not.
- 4) Alarm raising if necessary and further assessment – if the sound is classified to be that belonging to a road cutter, an alarm will be raised and a report will be sent to the control center so that further assessment can be made and measures taken.

Location of the potential threat can be well identified by knowing which sensor is raising an alarm. Thus when a sensor detects a road cutter, it will give off an alarm to caution the

people nearby of the underground pipeline. Meanwhile it will also send off a message to the control center to report the potential threat so relevant measures can be executed quickly. The whole monitoring process is depicted in Fig. 2.

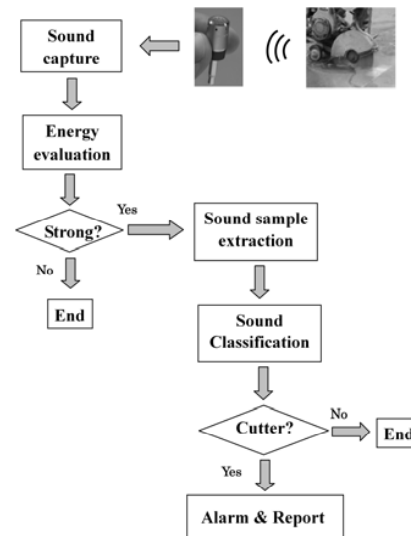


Fig. 2 Flow chart of the pipeline monitoring system

III. SOUND RECOGNITION

Although most acoustic recognition techniques were developed initially for speech recognition, other applications included environmental sound recognition mainly for the purpose of content-based classification, context awareness and ubiquitous surveillance.

For environment sound recognition, most of the popular research focuses on features of Mel Frequency Cepstral Coefficients (MFCC), Linear Prediction Coding Cepstrum (LPCC), etc., and on classifiers such as Euclidean distance, Vector Quantization (VQ), Support Vector Machine (SVM), Hidden Markov Models (HMM), Gaussian Mixture Model (GMM), k-nearest neighbor algorithm (KNN) and Neural Network (NN). Gaunard [8] classified five types of noise events using LPCC feature combined with a HMM classifier and showed a good result. Later, Peltonen [9] used MFCC and GMM to classify 10 inside and outside environments for scene recognition. Lu [7] classified five classes of sounds using MFCC and SVM in his audio segmentation and classification with good classification resolution achieved. In [10], Lu further pointed out that SVM is also much better than KNN and GMM. Toyoda [11] tried the multilayered Neural Networks for robotic audition. Krishna [12] compared MFCC and LPCC performance in musical instrument recognition and concluded that LPCC did better than MFCC. Ma [6] used the MFCC together with a HMM classifier to get a high resolution classification.

However, even though HMM showed a good performance, it usually requires large amount of training data to accurately train the models. Large computation cost makes it inconvenient

for the small sensors. Above all, HMM, VQ as well, is a method for multi class classification and it is hard to be applied to the one class classification problem which is the case in our road cutter recognition. Neural Network method and GMM also has the problem of having a large computation burden. Since SVM seems to be better than GMM and KNN [10], and could be much faster as the computation is only depends on small number of supporting vectors, a one class SVM classifier is used for our proposed system.

As for the features, MFCC and LPCC are cespectrum based features and a little too complicated. More over, they can only be used for recognizing a signal frame. In this paper, however, we would like to use a PCA-one class SVM based approach so that a several second sound sample with dozens of frames can be recognized all at once as a whole. This approach is introduced below in detail.

A. Mechanism of PCA and SVM based Sound Recognition

In most conventional sound recognition processes, a several second sound sample is often segmented out for testing. In the signal processing, it is usually further segmented into many tiny frames, usually lasting from 20ms to 30ms, with an overlap between every two adjacent frames. Acoustic recognition will then be carried out according to the features of all overlapped frames. Ma [6] recognized an environment sound based on a 3s sample, with 25ms frames and 15ms overlap. Goldhor [13] further pointed out that the overlap usually has to be more than 25% of the frame size. On the other hand, Lu [7] classified a 1s sample, with 40 evenly segmented non-overlapping 25ms frames. Statistical characteristics over all 40 frame features were used to classify the sounds. Although Lu's method required only short samples and no frame overlapping, the computing cost and memory requirement is still quite significant for tiny smart sensors. Also their approaches suffer the same problem, i.e., the last decision could not be made without features of all frames. This means every frame needs to be processed individually at first. The acquired data also need to be reserved. However, a sound sample usually contains many of such frames so that both the computation and memory cost will be huge.

In order to lower the computational cost and memory requirement of the sensors, as well as its energy consumption, in this paper, however, another approach using PCA and one class SVM is applied, which can make a decision for a sound sample as a whole. More over, some individual frames interfered by the noise will be avoided affecting the last decision, due to the denoising ability of the PCA. For further decreasing the computation cost, considering of the monotonous characteristic of the noise emitted from a road cutter in constant operation, we proposed a separated frame blocking mechanism [6]. For each sound sample, all frames are segmented separately. An interval is set between every two adjacent frames instead of the overlap. Then those separated frames would provide us their power spectral density (PSD), which will be processed by PCA. After that, a feature vector, truncated eigenvalues, could be obtained. It is obvious that the

acquired feature vector actually characterizes the whole sound sample, but rather than the individual frames. Based on the obtained feature, a one class SVM classifier will then be applied to make the classification and the decision can be made. The PCA and one class SVM based sound recognition process is briefly depicted in Fig. 3.

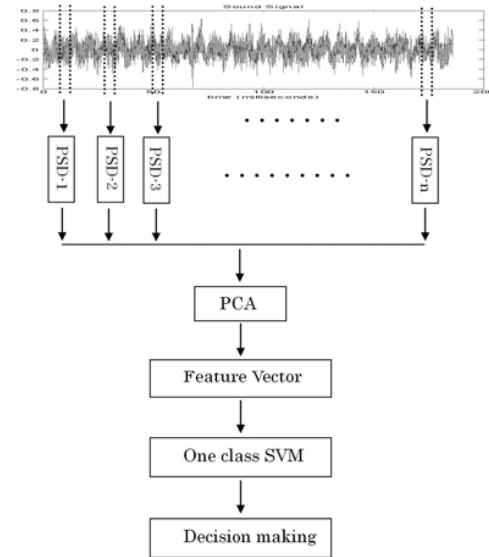


Fig. 3 PCA and one class SVM based sound recognition

B. Principal Component Analysis

From the incoming sound sample, n separated frames will be segmented out for analysis. For every frame, the signal is transformed into frequency domain by FFT transform, so that the power spectral density (PSD) could be obtained. For the following pattern comparison, PSD of each frame is normalized into unit power within all frequency range. Thus n normalized PSDs could be obtained as:

$$\mathbf{P}_i = (f_{i1}, f_{i2} \cdots f_{ik}) \quad (i = 1, 2, \cdots, n) \quad (1)$$

where: \mathbf{P}_i is PSD vector of the i th sound frame; k is the FFT index; and $\sum_{i=1}^k f_{ij} = 1$.

The covariance matrix \mathbf{C} can be expressed as:

$$\mathbf{C} = E\{(\mathbf{P} - E(\mathbf{P}))(\mathbf{P} - E(\mathbf{P}))^T\} \quad (2)$$

From this symmetric covariance matrix, a unit orthogonal basis can be obtained by finding its eigenvalues and eigenvectors. By setting the eigenvalues in a descending mode, the first eigenvector will point to the direction of largest variance of the data, while the second eigenvector has the direction of second largest variance and so on. The eigenvalue vector \mathbf{A} can be expressed as:

$$\mathbf{A} = (a_1, a_2, \cdots, a_n) \quad (3)$$

where:

$$a_1 \geq a_2 \geq a_3 \cdots \geq a_n \quad (4)$$

In PCA analysis, usually only first several principal components are needed. A truncated eigenvalue vector $\tilde{\mathbf{A}}$ is used for our feature vector which could be written as:

$$\tilde{\mathbf{A}} = (a_1, a_2, \dots, a_m) \quad m < n \quad (5)$$

Every kind of sound has its own pattern and thus has its own principal space. Eigen vectors are actually the direction vectors, while eigenvalues represent the weights for each direction. In most cases, the truncated eigenvalues could uniquely characterize the sound. Thus using truncated eigenvalues $\tilde{\mathbf{A}}$ as the feature vector is feasible, practical and reasonable.

C. One Class SVM

Support Vector Machines (SVM) is a supervised learning method which can separate the data easily using a hyperplane by projecting them into a higher dimensional feature space. SVM was first introduced by Vapnik et al. and soon became popular due to its strong power of classification and many successful applications.

SVM was initially used for classifying two classes. But it was soon extended to the use for multi-class and one class classification. For one class classification problems, they are often caused by lack of data or incomplete information. In our road cutter recognition, considering that it is impossible for us to collect all kinds of environmental sounds to train them, it is reasonable to apply one class SVM to make classification.

The one class SVM classifier distinguishes other classes from a known class, depending on the decision hyperplane built on the support vectors and a Kernel function. The decision function for one class SVM has the form [14]:

$$f(x) = \text{sgn}\left(\sum_{i=1}^l \alpha_i \cdot K(\Phi(X_i), \Phi(Z)) - \rho\right) \quad (6)$$

where: $\Phi(X_i)$ is a support vector, $\Phi(Z)$ is a data vector to be classified, K is a Kernel function, l is the number of the support vectors and ρ is the offset.

IV. EXPERIMENTS AND RESULTS

A. Experiment Overview



Fig. 4 An experiment was conducted at Tokyo

In order to test the feasibility of deploying this monitoring system to a real pipeline, extensive experiments were conducted. The experiments were conducted at several places

in Tokyo at different times. Fig. 4 shows an experiment conducted at Tokyo. During the experiments, a microphone (Sony ECM-CR120) was used to capture the sound, and a digital recorder (Olympus IC recorder) to record it. The specification of the experiment equipments can be shown in Table I.

TABLE I
EXPERIMENTAL EQUIPMENT SPECIFICATION

Microphone		Digital Recorder	
Model	ECM-CR120	Model	Olympus V-60
Directivity	Omni-directional	Sampling frequency	44.1KHz
Response range	100Hz-12KHz	Response range	50Hz-13KHz
Sensitivity	-46dB±4dB	Input level	-70dBv

We collected 151 sound samples in total, including road cutting sounds and other environmental sounds. As shown in Table II, we picked up 10 road cutting sounds as the cutting template sounds which were used to build the one class support vector machine. 15 other sounds, including 5 cutting sounds and 10 non-cutting sounds, were used for the training sounds, to train and optimize the Gaussian kernel parameters of the one class SVM, such as highest allowable fraction of the misclassification and the bandwidth of the Gaussian distribution. The rest of 126 sounds were used for testing.

TABLE II
EXPERIMENT DATA COLLECTION

	Total sounds	Template sounds	Training sounds	Testing sounds
Road cutter	51	10	5	36
Vehicle	15		1	14
Backhoe	44		7	37
Train	4			4
Wood cutting	7			7
Pionjar drill	5			5
Others	25		2	23
Total	151	10	15	126

For analyzing the sound, we studied the power spectra of the road cutter samples at first. Usually the sound contains two parts, the low frequency band sound from the engine of the road cutter and the high frequency band sound from the action of road cutting. It can be found that the frequencies vary significantly in different samples even though they all belong to the same class, i.e., road cutter, as shown in Fig. 5. The variation of the frequency can be explained by the working conditions. The material the blade is cutting is always changing, either the soft asphalt, or sand, or the hard carpolite or something else. Also the pressure put on the cutter by the worker can never be kept constant. The lubricant effect of the

water, which is indispensable when the cutter is working, will change the frequency too. Thickness of the road also results in the frequency variation. At last, another fact we need to treat seriously is that the engine condition of different cutters varies significantly too. Usually, for a new cutter the low frequency component will be small while for an old cutter, the low frequency engine sound may be very large, making high frequency cutting sounds relatively trivial. All of these facts cause big problems for generalizing the classification process.

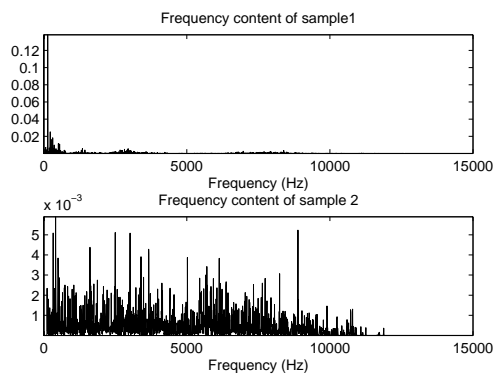


Fig. 5 Significant frequency variation of cutter sounds

Since most environmental noise sounds from the street have relatively low frequency components, such as engine sounds of vehicles and voices of pedestrians, etc., interference may easily happen in the low frequency band. Moreover, the actual danger to the pipeline is coming from the action of road cutting. It seems that the road cutting sound should be the deterministic sound that actually characterizes a road cutter. In this sense, it seems that in order to effectively recognize a dangerous road cutter, its engine sound and road cutting sound should be separated. As we found that the frequencies from the cutter engine were usually around 140Hz, the effect for frequencies below 500Hz being either remained or removed were both studied in our research.

B. PCA Analysis and One Class SVM Classification

For each sound sample, 25 tiny frames were segmented out, each lasting 23ms with a 200ms interval between every two adjacent frames. The normalized PSD were further analyzed by PCA. Thus the feature vector, eigenvalue vector, was obtained. Fig. 6 and Fig. 7 show the eigenvalues (without being truncated) and the accumulated proportion of variance for 10 template cutter sounds in both conditions when low frequency engine sound was remained and removed respectively.

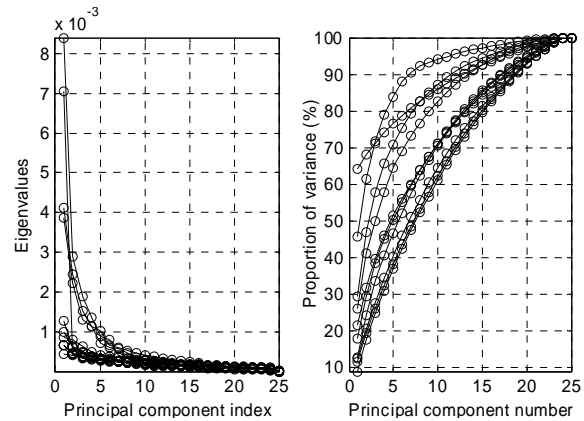


Fig. 6 Eigenvalues and proportions of variance for low frequency engine sound being remained

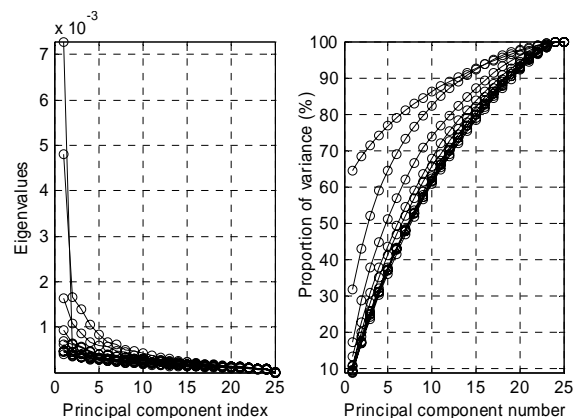


Fig. 7 Eigenvalues and proportions of variance for low frequency engine sound being removed

For principal component analysis, an important issue is to decide how many principal components (PCs) we should keep, i.e., decide the dimension of the principal space. Usually, it could be decided by following some criterions such as the Kaiser criterion, the variance proportion criterion which finds the number of components that comprise some part of the total variance, and the scree test criterion, which suggests to find the place where the smooth decrease of eigenvalues appears [15] [16][17]. It was pointed out that the Kaiser criterion may lead to too many PCs being taken while the scree test may lead to too few PCs. For simplicity, we focused on the last two criterions in this paper. From the eigenvalues in Fig. 6 and Fig. 7, according to the scree test criterion, the number of PCs should be around 5. However, according to proportion of variance, also shown in Fig. 6 and Fig. 7, about 15 PCs are needed even for 80% of total variance. Thus the PCs should be between 5 and 15. Also, due to the significant eigen value variation caused by frequency variation between samples, it is very difficult for us to decide the exact dimension of the principal space for our case. In this paper, we therefore tested 3 conditions, when the number of PCs is set to be 5, 10, and 15.

After the truncation of the eigenvalue serials, the remained eigenvalues were used for the feature vector. The feature vector was then be classified by the trained one class SVM classifier. Usually, the hyperplane of the one class SVM is the geometric form of the decision function. When the discriminant value is larger than 0, the testing signal could be regarded as in the same class, otherwise it would be regarded as in different class. However, for this pipeline monitoring, considering the ponderance of the accidents, we would like to decrease the rejection error (regarding it is not a road cutter but actually it is), even though it would increase the risk of acceptance error (regarding it is a road cutter but actually it is not). By studying the discriminant values of the training samples, as shown in Fig. 8, it could also be found that most of the non-cutter sound were well recognized so that their distances to the hyperplane were very large, thus leave us large space for adjusting. For these considerations, the hyperplane was deliberately shifted a little away from the road cutter class, to be -0.01, so that the margin for cutter class would be increased and could tolerate more suspicious samples.

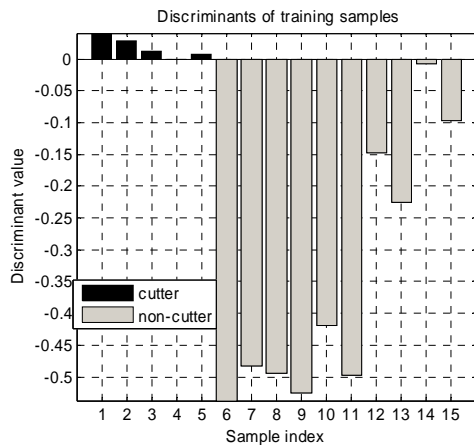


Fig. 8 Discriminant values for training samples with PCs being 10 and engine sound remained

By this way, all the testing sounds were classified and the recognition decision errors were listed in the Table III. Results show that the PCA and one class SVM based cutter recognition algorithm can do the work very well. With low frequency engine sound remained and PCs being 10, only 1 cutter sound was incorrectly classified, while all the other sounds were recognized successfully and correctly. The recognition algorithm also works well when engine sound was removed, with totally 4 sounds being mistakenly recognized. The overall success rate, i.e., the success rate of either road cutter sound being correctly recognized as road cutter or non-cutter sound being correctly recognized as not from a road cutter, reaches as high as 99.21% when low frequency engine sound was remained, and 96.83% when engine sound was removed. Even though it seems that separating engine sound and road cutting sound could improve the recognition correctness, our experiments, however, showed an opposite result. The success rate when low frequency engine sound was remained is slightly better than that when the engine sound was removed. This is due to the robustness of the SVM. By the one class SVM, the engine sound was also helped to make classification in some extent. From the table we can also find that principal space dimension of 5 can basically satisfy our needs. Principal space dimension being 10 works best when the engine sound was remained.

V. CONCLUSION

In this paper, acoustic information is used to recognize dangerous construction machines that are potential threats to the well-being of underground pipelines. An automatic pipeline monitoring mechanism is proposed. A sound recognition approach based on PCA and one class SVM was studied and applied. With this approach, a sound sample can be recognized as a whole, instead of making decision based on many initially recognized frames. At the same time, abstracting ability of the PCA makes it robust to noises. Real site experiments were conducted and data were analyzed. Results showed that PCA and one class SVM based algorithm can do the work very well.

TABLE III
RECOGNITION ERROR

Test sounds		Engine sound remained			Engine sound removed		
		dim=5	dim=10	dim=15	dim=5	dim=10	dim=15
Cutter	36	1	1	1	1	1	1
Vehicle	14	1	0	0	0	0	0
Backhoe	37	2	0	0	0	0	0
Train	4	0	0	0	0	0	0
Wood cutting	7	0	0	0	1	1	1
Pionjar drill	5	0	0	0	2	2	2
Others	23	0	0	0	0	0	0
Total	126	4	1	1	4	4	4

dim stands for the dimension of the principal space applied, i.e., the number of PCs retained.

The automatic pipeline monitoring system and sound recognition technologies studied in this paper will be very useful for pipeline monitoring sensor network systems in the future to prevent potential damage and ensure the safety of underground lifeline infrastructures.

ACKNOWLEDGMENT

We would like to thank Mr. T. Kume and Dr. M. Tamura at Tokyo Gas Co., Ltd. for providing valuable experimental data.

REFERENCES

- [1] R.J. Eiber, R.J.; D.J. Jones, G.S.Kramer, "Outside force causes most natural gas pipeline failures", *Oil and Gas Journal*, vol.85, issue 11, pp.52-57, March 1987.
- [2] D. Hausamann, et.al, "Monitoring of gas transmission pipelines-A customer driven civil UAV application", ODAS Conference, 2003.
- [3] JE. Huebler, "Detection of Unauthorized Construction Equipment in Pipeline Right-of-Ways", Technical Report of Gas Technology Institute, 2004.
- [4] C. Wan, A. Mita and T.Kume, "An automatic pipeline monitoring system using sound information", *Structural Control and Health Monitoring*, to be published ; published inially within the online version of the *Structual Control and Health Monitoring*. DOI : 10.1002/stc.295.
- [5] C. Wan and A. Mita, "Recognition of potential danger to buried pipelines based on sounds ", *Structural Control and Health Monitoring*, to be published.
- [6] L. Ma, B. Milner and D. Smith, "Acoustic environment classification", *ACM TSLP*, vol.3, issue.2, pp.1-22, July 2006.
- [7] L. Lu, S.Z. Li, H.J. Zhang, "Content-based audio segmentation using support vector machines", *IEEE international conference on Multimedia and Expo*, pp.956-959, 2001.
- [8] P. Gaunard, C.G. Mubikangiey, C. Couvreur, V. Fontaine, "Automatic classification of environmental noise events by hiddenMarkov model", *Proc. IEEE*, vol.6, pp.3609-3612, May 1998.
- [9] V. Peltonen. J. Tuomi, A. Klapuri, J. Huopaniemi and T. Sorsa, "Computational auditory scene recognition", *Proc. IEEE*, vol.2, pp.1941-1944, May 2002.
- [10] L. Lu, S.Z. Li, H.J. Zhang, "Content-based audio classification and segmentation by using support vector machines", *Multimedia Systems*, vol.8, no.3, pp.482-492, 2003.
- [11] Y.Toyoda, J. Huang, S. Ding and Y. Liu, "Environmental sound recognition by multilayered neural networks", *Proc. IEEE, CIT*, pp.123-127, Sep 2004.
- [12] A.G. Krishna and T.V. Sreenivas, "Music instrument recognition: from isolated notes to solo phrases", *Proc. IEEE*, vol.4, pp.265-268, May 2004.
- [13] R.S. Goldhor, "Recognition of Environment Sounds", *Proc.IEEE on Acoustics, Speech, and Signal Processing*, Vol.1, pp.149-152, Apr 1993.
- [14] R. Unnthorsson, T.P. Runarsson and M.T. Jonsson, "Model selection in one class nu-SVMs using RBF kernels", *16th Conference on Condition Monitoring and Diagnostic*, pp.1-11, April 2003.
- [15] Principal Component analysis (PCA) or Empirical Orthogonal Function (EOF), Lecture notes, Lunds University. Available: <http://aqua.tvrl.lth.se/course/VVR005F/2%20EOF.pdf>
- [16] Principal Component analysis, notes from Indiana University, Available: <http://cheminfo.informatics.indiana.edu/~rguha/writing/notes/stats/node7.html>
- [17] Principal Components and Factor Analysis, electronic statistics textbook, StatSoft, Inc, Available : <http://www.statsoft.com/textbook/stfacan.html>