

Document downloaded from the institutional repository of the University of Alcalá: <http://ebuah.uah.es/dspace/>

This is a postprint version of the following published document:

Tejedor, J., Macías-Guarasa, J., Martins, H. F., Martín-López, S. & González-Herráez, M. 2018, "A Gaussian Mixture Model-Hidden Markov Model (GMM-HMM)-based fiber optic surveillance system for pipeline integrity threat detection", in 26th International Conference on Optical Fiber Sensors, OSA Technical Digest (Optical Society of America, 2018), paper WF36

Available at <http://dx.doi.org/10.1364/OFS.2018.WF36>

© 2018 Optical Society of America. Users may use, reuse, and build upon the article, or use the article for text or data mining, so long as such uses are for non-commercial purposes and appropriate attribution is maintained. All other rights are reserved.

(Article begins on next page)



This work is licensed under a

Creative Commons Attribution-NonCommercial-NoDerivatives
4.0 International License.

A Gaussian Mixture Model-Hidden Markov Model (GMM-HMM)-based fiber optic surveillance system for pipeline integrity threat detection

J. Tejedor (1), J. Macias-Guarasa (2), H. F. Martins (2), S. Martin-Lopez (2), M. Gonzalez-Herraez (2)

(1) Fundación Universitaria San Pablo CEU, Madrid, Spain, (2) Universidad de Alcalá, Alcalá, Madrid. Spain
javier.tejedornogueras@ceu.es

Abstract: A pipeline integrity threat detection system using Distributed Acoustic Sensing and Artificial Intelligence (AI) is presented. The AI uses a combination of Gaussian Mixture Models and Hidden Markov Models (GMMs-HMMs), outperforming our former GMM-based system.

OCIS codes: 060.0060 General; 060.2370 Fiber optics sensors; 070.5010 Pattern Recognition

1. INTRODUCTION

Pipeline transmission is the most sustainable and safest transmission method to transport energy from the production facilities to the end-users. Special attention must be paid by the system transmission operators aiming to avoid a real damage [1,2]. Distributed Acoustic Sensing (DAS) is a suitable technology to address this task [3-5] since it has the sensitivity to detect potential threats for the pipeline integrity so that the corresponding corrective action can be carried out. However, for the DAS system to be fully operational, a Pattern Recognition System (PRS) is needed to classify each of the potential activities into potential threats or non-threats, hence reducing the number of false alarms in the system to an acceptable level and increasing the cost-effectiveness of the solution.

This paper presents an improved version of a real-time, smart fiber optic surveillance system based on DAS+PRS aiming to classify certain activities and detect potential threats that occur in the pipeline vicinity. This system is a combination of hardware and software modules. The hardware module continuously monitors vibrations in an optical fiber along a gas pipeline and is based on a phase-sensitive optical time domain reflectometer (Φ -OTDR) sensor named Fiber Network Distributed Acoustic Sensor (FINDAS) [6]. The software module is an original PRS based on Gaussian Mixture Models-Hidden Markov Models (GMMs-HMMs) that operates in two modes: 1) The machine+activity identification mode classifies different activities occurring in different locations along the pipeline, and 2) the threat detection mode aims to detect threats that occur along the pipeline.

A GMM-HMM-based system was previously proposed to detect fence intrusion vibration signals [7]. However, that work employed Sagnac interferometers for signal acquisition (not a DAS), and only two classes were experimented with (intrusion and non-intrusion). In addition, the signal recordings comprised rattling and sweeping on the optical fiber cables manually, which does not correspond to a real-field deployment configuration, and no information related to the locations of the recordings was given. To the best of our knowledge, this is the first report that employs GMMs-HMMs in a DAS+PRS-based surveillance system for real-time monitorization of long pipelines. Our work addresses two different configurations (machine+activity identification and threat detection) and is evaluated on real field data acquired in various scenarios, being tested using a rigorous evaluation procedure.

2. DISTRIBUTED ACOUSTIC SENSING SYSTEM

The DAS system consists of a standard Φ -OTDR sensor called FINDAS [6,8]. The FINDAS sensor has an (optical) spatial resolution of 5m (readout resolution of 1m) and a typical sensing range of up to 45km, using standard single-mode fiber (SMF). The FINDAS sensor was connected to a standard SMF, which had been previously installed parallel along the pipeline. Since the fiber does not always follow on a parallel with the pipeline and, for maintenance purposes, there were fiber rolls in some locations, the correspondence between fiber distance and pipe length is not consistent. Therefore, a calibration between fiber distance and field location was done before the tests, by applying a known signal at a given field location and matching it with the fiber position of the recorded signal.

If the energy of the vibrations monitored by the FINDAS is higher than a given energy threshold at a certain location, an alarm was raised by the FINDAS, stating that there was activity at that location. At this point, the acoustic samples were recorded and sent to the pipeline integrity threat detection system (described in Section 3), which evaluated if the activity was considered a threat or a non-threat for the integrity of the pipeline and addressed machine+activity identification as well. If the activity was considered a threat for the integrity of the pipeline, an alarm was sent to the end user. The FINDAS energy threshold alarm level is adjustable for each

location, thus allowing to consider factors such as background noise and fiber sensitivity. The simultaneous detection of multiple activities at different locations is also possible. The system architecture is shown in Fig. 1.

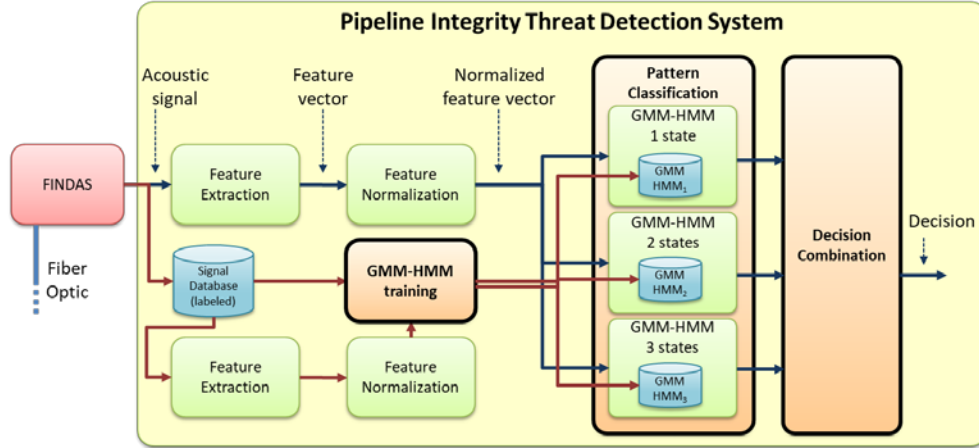


Figure 1. Surveillance system architecture.

3. PIPELINE INTEGRITY THREAT DETECTION SYSTEM

The pipeline integrity threat detection system consists of four different modules: the *feature extraction* module, which extracts the most relevant information from the acoustic traces recorded by the FINDAS, the *feature normalization* module, which normalizes the features to cope with the strong variability between the acoustic signals recorded and the sensed positions, the *classification* module, which bases on pattern recognition and assigns the acoustic trace three decisions by comparison with a set of previously trained patterns, according to a certain algorithm, and the *decision combination* module, which combines the three decisions output by the classification module into a single class (machine+activity or threat/non-threat depending on the system mode).

The feature extraction is based on the spectral content of the acoustic traces, which are analyzed by sequentially splitting them in acoustic frames of a given length. This spectral information is computed using the Short Time Fast Fourier Transform (ST-FFT), and is then integrated as the energy of a set of frequency bands. The values of the spectral energy in bands constitute the final feature vector components that represent each acoustic frame of the input signal. The relevant parameters for the ST-FFT are the acoustic frame size (set to 1 second), the acoustic frame overlap (set to 5 milliseconds), and the number of FFT points (set to 8192). The number of frequency bands defines the number of components in the feature vectors, and is set to 100. All these figures were selected due to their best performance in preliminary experiments. As feature normalization, the sensitivity-based normalization employed in [5] was used.

The classification is based on Gaussian Mixture Models-Hidden Markov Models [9] and comprises two different processes: training and recognition, which are explained below.

3.1 GMM-HMM training

The training process is carried out with data recorded in several field test recordings, and just needs to run once. This builds a GMM-HMM with a single Gaussian component for each of the activities in the machine+activity identification mode, and two different GMM-HMMs (one representing the *threat* class and the other representing the *non-threat* class) with a single Gaussian component each, in the threat detection mode. The GMM-HMM training consists of the estimation of the mean and the full covariance matrix of the Gaussian component, and the transition matrix probabilities for each state of the HMM [9]. Fig. 2 shows an example of a 3-state HMM, where each state in the HMM is modelled with a single Gaussian.

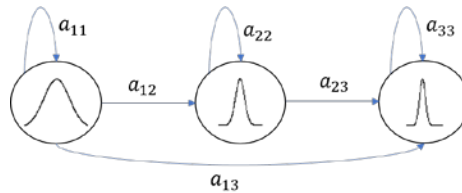


Figure 2. Hidden Markov Model with 3 states. The symbols a_{ij} denote the transition matrix probabilities.

To cope with the different behavior of the activities to be detected, three training processes were run, which only differ in the number of states used in the model (from 1 to 3 states per HMM). These three sets of GMM-HMMs are then used in the recognition stage.

3.2 GMM-HMM recognition

Viterbi algorithm [9] is used to classify each test acoustic frame as the class (machine+activity or threat/non-threat) with the highest probability. The Viterbi algorithm finds the “best” path between the test acoustic frames and the GMM-HMMs previously trained. Three recognition processes (each using the set of GMM-HMM with 1, 2, and 3 states) are run to compute three individual frame-level decisions for each test acoustic frame.

3.3 Decision combination

From the three decisions that comprise the output of the classification step, the class that is finally assigned to each test acoustic frame is chosen by majority voting, so that the selected class is the one that obtains more votes given the three recognition processes.

4. FIELD TESTS AND RESULTS

A gas transmission pipeline operated by Fluxys Belgium S. A. was used for the field tests, thus aiming to operate in a real case scenario. Activities of different machinery were recorded near the pipeline by monitoring an optical fiber cable installed at about 0.5m away from the pipeline and parallel to it. The FINDAS was placed in a telecom room, and activity was recorded at six different locations at an (optical fiber) distance from the FINDAS of 22450m, 22700m, 23950m, 27650m, 27800m, and 34500m.

Four different machines, as shown in Fig. 3, which carried out different activities, were recorded in those locations to create the 10-hour acoustic database used to test the surveillance system. The database, which was created with acoustic signals recorded at a sample frequency of 1085Hz, consists of 8 different machine+activity pairs, as follows: a 5ton Kubota KX161-3 (moving, hitting the ground, scrapping the ground); a 1,5ton Kubota KX41-3V (moving, hitting the ground, scraping the ground); a pneumatic hammer (hitting the ground); and a plate compactor (compacting the ground). These activities were then assigned the *threat* or *non-threat* class for the threat detection mode of the system.



Figure 3. Machines used for the fields tests: a) Kubota KX161-3 (~5000kg; 5.54x2.53x1.96m length/height/width), b) Kubota KX41-3V (~1500kg; 3.66x2.28x0.99m length/height/width), c) Pneumatic hammer, d) Plate compactor

The tests were conducted by cross-validation (CV) on the recorded data in a frame-basis. This approach is required to obtain a reliable estimation of the actual classification accuracy. The evaluation uses a six-fold CV, where each fold contains the data from five locations for the GMM-HMM training, and the data belonging to the other location are used for actual testing. The final results are calculated by averaging the performance of each of the six-fold results, according to the frame-based classification strategy.

Real/Recognize	1	2	3	4	5	6	7	8
1 2.a) moving (NT)	44.5%	22.9%	13.7%	7.1%	2.6%	9.1%	0.2%	0.0%
2 2.a) hitting (T)	11.3%	35.9%	15.1%	1.9%	9.4%	24.5%	0.0%	1.9%
3 2.a) scrapp. (T)	1.9%	16.7%	31.5%	4.3%	16.7%	27.8%	0.0%	1.2%
4 2.b) moving (NT)	14.4%	9.8%	5.9%	46.6%	2.6%	19.0%	1.0%	0.7%
5 2.b) hitting (T)	5.3%	5.3%	7.0%	7.0%	31.6%	36.8%	0.0%	7.0%
6 2.b) scrapp. (T)	3.9%	0.8%	8.5%	6.9%	22.3%	53.1%	0.0%	4.6%
7 2.c) hitting (NT)	9.3%	0.5%	0.0%	0.5%	1.0%	2.6%	78.4%	7.7%
8 2.d) compact. (NT)	5.0%	1.4%	0.0%	9.1%	28.2%	19.1%	5.9%	31.4%

Table 1. Confusion matrix for the GMM-HMM-based system. In rows, the real class is presented. In columns, the recognized class output by the GMM-HMM-based system is presented. 2.a), 2.b), 2.c), and 2.d) denote the machines in Figure 2. (NT) represents “non-threat” activities, and (T) represent “threat” activities.

Real/Recognize	1	2	3	4	5	6	7	8
1 2.a) moving (NT)	49.1%	13.8%	12.8%	9.3%	5.7%	6.4%	1.4%	1.6%
2 2.a) hitting (T)	20.0%	20.1%	20.4%	6.1%	8.2%	15.6%	1.7%	7.8%
3 2.a) scrapp. (T)	11.9%	12.1%	26.0%	7.6%	11.8%	20.4%	1.2%	9.1%
4 2.b) moving (NT)	13.2%	5.8%	10.1%	50.5%	5.6%	9.4%	3.4%	2.0%
5 2.b) hitting (T)	10.7%	7.0%	15.4%	6.5%	13.8%	28.0%	2.5%	16.1%
6 2.b) scrapp. (T)	6.6%	6.9%	16.8%	6.6%	14.3%	30.2%	2.6%	16.1%
7 2.c) hitting (NT)	5.9%	1.7%	1.8%	4.3%	2.9%	5.2%	71.8%	6.5%
8 2.d) compact. (NT)	6.5%	3.8%	6.9%	4.9%	10.5%	22.4%	5.5%	39.5%

Table 2. Confusion matrix for the GMM-based system. The layout is the same as in Table 1.

To compare our proposed GMM-HMM-based system with a GMM-based baseline approach [5], the confusion matrix of each approach is presented in Tables 1 and 2, respectively. The GMM-HMM-based system obtains much better performance for the classes that perform worse in the baseline GMM-based approach (classes 2, 3, 5, and 6), with a slight degradation in the classes that perform better in the baseline approach (classes 1, 4 and 8). The average classification accuracy is of 45.68% in the machine+activity identification mode, which represents some improvement over the 45.15% baseline accuracy for the GMM approach. The average effect is not high since the classes that perform the best in the GMM-based approach have a higher presence in the recorded database, but the large improvement in the classes that perform worse allows to assess that the GMM-HMM-based system is worthy.

For the threat detection mode of the system, the GMM-HMM-based system presents a 91% of threat detection, with a 53.7% of false alarms. On the other hand, the GMM-based system presents an 80% of threat detection and a 40% of false alarms. This indicates that the GMM-HMM-based system is more suitable for scenarios in which a miss (i.e., a real threat that is not output by the system) is more relevant than generating a false alarm.

5. CONCLUSIONS

This paper has presented a novel DAS+PRS-based surveillance system that aims to detect potential threats in a gas pipeline and that also addresses machine+activity classification. The PRS is based on Gaussian Mixture Models-Hidden Markov Models. Results have shown that the GMM-HMM-based system outperforms a GMM-based system for machine+activity classification. Specifically, the GMM-HMM-based system gives more benefit to the “most complex” activities such as those involving more than one behavior (hitting/scrapping) than to the “easier” and more regular activities (movement), for which a *simpler* GMM-based system performs better. Regarding threat detection, the GMM-based system presents better overall results, while the GMM-HMM-based approach proved to be more suitable for scenarios in which a miss threat is more relevant than generating a false alarm.

ACKNOWLEDGEMENTS

Authors were supported by funding from the European Research Council through Starting Grant UFINE (Grant Number #307441), Water JPI, the WaterWorks2014 Co-funded Call, the European Commission (Horizon 2020) through Project H2020-MSCA-ITN-2016/722509-FINESSE, the Spanish Ministry of Economy and Competitiveness, the Spanish “Plan Nacional de I+D+I” through Projects TEC2015-71127-C2-2-R, TIN2013-47630-C2-1-R and TIN2016-75982-C2-1-R, by the University of Alcala through project CCGP2017-EXP/025, and by the regional program SINFOTONCM: S2013/MIT-2790, funded by the “Comunidad de Madrid”. HFM acknowledges funding from the FP7 ITN ICONE program (Grant #608099) funded by the E. C., and SML was supported by funding from the Spanish Ministry of Science and Innovation through a “Ramón y Cajal” contract.

REFERENCES

- [1] P. S. Trust. “Nationwide data on reported incidents by pipeline gas (gas transmission, gas distribution, hazardous liquid,” <http://pstrust.org/about-pipelines1/stats/accident/> [Last access: March 2018].
- [2] E. G. P. I. D. G. (EGIG). “9th egig-report on gas pipeline incidents (1970-2013),” <https://www.egig.eu/reports>. Tech. Rep., 2015. [Last access: March 2018].
- [3] Q. Sun, H. Feng, X. Yan, Z. Zeng. “Recognition of a phase-sensitive OTDR sensing system based on morphologic feature extraction,” *Sensors*, 15, 15179-15197 (2015).
- [4] H. Wu, S. Xiao, X. Li, Z. Wang, J. Xu, Y. Rao, “Separation and determination of the disturbing signals in phase-sensitive optical time domain reflectometry (Φ -OTDR),” *Journal of Lightwave Technology*, 33(15), 3156-3162 (2015)
- [5] J. Tejedor, H. F. Martins, D. Piote, J. Macias-Guarasa, J. Pastor-Graells, S. Martin-Lopez, P. Corredera, F. D. Smet, W. Postvoll, M. Gonzalez-Herraez. “Towards prevention of pipeline integrity threats using a smart fiber optic surveillance system,” *Journal of Lightwave Technology*, 34(19), 4445-4453 (2016).
- [6] H. F. Martins, S. Martin-Lopez, P. Corredera, M. L. Filograno, O. Frazao, M. Gonzalez-Herraez. “Coherent noise reduction in high visibility phase sensitive optical time domain reflectometer for distributed sensing of ultrasonic waves,” *Journal of Lightwave Technology*, 31(23), 3631-3637 (2013).
- [7] Z. Ma, Y. Liang, J. Zhu. “An optic-fiber fence intrusion recognition system using mixture Gaussian hidden Markov models,” *IEICE Electronics Express*, 14(5), 1-8 (2017).
- [8] H. F. Martins, S. Martin-Lopez, P. Corredera, M. L. Filograno, O. Frazao, M. Gonzalez-Herraez. “Phase-sensitive optical time domain reflectometer assisted by first-order Raman amplification for distributed vibration sensing over > 100km,” *Journal of Lightwave Technology*, 32(8), 1510-1518 (2014).
- [9] L. R. Rabiner. “A tutorial on Hidden Markov Models and selected applications in speech recognition,” *Proc. of the IEEE*, 77(2), 257-286 (1987).