Abstract

Oil and gas pipeline leakages lead to not only enormous economic loss but also environmental disasters. How to detect the pipeline damages including leakages and cracks has attracted much research attention. One of the promising leakage detection method is to use lead zirconate titanate (PZT) transducers to detect the negative pressure wave when leakage occurs. PZT transducers can generate and detect guided stress waves for crack detection also. However, the negative pressure waves or guided stress waves may not be easily detected with environmental interference, e.g., the oil and gas pipelines in offshore environment. In this paper, a Gaussian mixture model based hidden Markov model (GMM-HMM) method is proposed to detect the pipeline leakage and crack depth in changing environment and time-varying operational conditions. Leakages in different sections or crack depths are considered as different states in hidden Markov models (HMM). Laboratory experiments show that the GMM-HMM method can recognize the crack depth and leakage of pipeline such as whether there is a leakage, where the leakage is.

*Contact author, xuemin.chen@tsu.edu
1 Introduction

Thousands of miles of pipelines crisscrossed on the Gulf of Mexico seafloor are the veins for offshore oil and gas industry of the U.S. or even the whole world, while the leaks and ruptures of those pipelines lead to not only enormous economic loss but also environmental disasters [1]. In the last few decades, many pipeline structural health monitoring techniques have been used to monitor damages [1, 2]. One of the promising methods for pipeline leakage detection is based on the negative pressure wave (NPW) [3–7]. NPW is generated at the leak point when the fluid or gas escapes in the form of a high velocity jet [8]. Then the NPW propagates along pipeline in both directions, i.e., the upstream and downstream of leakage point. The NPW can be detected by Lead zirconate titanate (PZT) transducers. PZT transducers are made of piezoelectric materials which can convert mechanical energy to electrical energy and vice versa. This piezoelectric effect leads to PZT transducers work as passive sensors or active actuators. PZT transducers can be effectively used as passive sensors to catch the acoustic signals propagating along the pipeline. On the other hand, the pipeline health condition needs to be evaluated periodically to provide early warning. To address this demand, PZT transducers have been used in active sensing mode to detect crack [9]. PZT transducers have found lots of applications for structural health monitoring. Wang et al. [10] invented a wearable PZT transducers which can be easily and noninvasively worn onto the flanged valve for bolted joint in real-time. Gong et al. [11] developed an algorithm to process signals collected by PZT sensor for automatic extraction of the stress wave reflection period. In these applications, the structural health condition is evaluated according to extracted features. However, in the actual situation, the changing environment and time-varying operational conditions make the reliability of damage evaluation facing the challenge. Especially for the offshore pipelines, the submarine environment is more complicated. With the development of machine learning, many attention has been paid to the probabilistic and statistical model-based methods, which are effective tools for characterizing uncertainties of signals such as Gaussian mixture model (GMM) [12], hidden Markov model (HMM) [13]. These researches have shown the potential to improve the damage evaluation reliability under uncertainties.

Among the existing probabilistic and statistical models [12, 13], HMM has a strong capability in pattern classification, especially for signals with nonstationary natures and poor repeatability and reproducibility [14]. HMM and
its variants have been extensively used for speech recognition [15], hand gesture recognition [16], handwritten word recognition [17], and newly applied to spam SMS detection [18]. In the field of structural health monitoring, several researchers have also applied the HMM to damage evaluation. Ram-mohan and Taha [19] using a standard HMM to model the simulated data of a pre-stressed concrete bridge. Tschpe and Wolff [20] studied the HMM for damage degree classification on plate-like structures. These studies indicate that the HMM is robust to uncertainties. So far, however, there is very little published research on using HMM detect pipeline leakage.

2 GMM-HMM based leakage detection

In this section, method for damage indexes extraction from the acoustic signals detected by PZT sensors is discussed. Then the design and training of GMM-HMM are presented.

2.1 Damage indexes extraction

To establish the relationship between characteristics change of the sampled waveform and damage parameters, two damage indexes are adopted to indicate the signal variations and serve as observations of the HMM model. The first damage index \(DI_1\) is a time-domain damage index, defined in [21]:

\[
DI_1 = 1 - \sqrt{\frac{\int (s_1(t) - \bar{s}_1)(s_2(t) - \bar{s}_2) \, dt}{\int (s_1(t) - \bar{s}_1)^2 \, dt \int (s_2(t) - \bar{s}_2)^2 \, dt}}.
\]

(1)

where \(s_1(t)\) is the baseline waveform and \(s_2(t)\) is the comparison waveform at time \(t\). The \(\bar{s}_1\) and \(\bar{s}_2\) is average value of \(s_1(t)\) and \(s_2(t)\). The baseline waveform represent the incident waveforms, and the comparison waveforms denote the captured waves detected by sensors. Unity minus the absolute value of the Pearson correlation coefficient is used as the time-domain damage index which can identify the signal difference.

The second one is a frequency-domain damage index \(DI_2\), the amplitude of peak frequency, as defined in [22]:

\[
DI_2 = \max_{f_1 \leq f \leq f_2} (|X(f)|),
\]

(2)

where \(X(f) = \int_{t_1}^{t_2} X(t) e^{-j2\pi ft} \, dt\). \(f_1\) and \(f_2\) are the start and stop frequency corresponding to the selected frequency spectrum window.
2.2 HMM and the BaumWelch algorithm

To reduce the calculation complexity and keep the model works efficiently, only first-order HMM is adopted in this study. A typical HMM model can be defined by a three-tuple:

$$\lambda = \{\pi, A, B\},$$

(3)

where $\pi$ is the initial probability distribution, $A$ is the state transition probability matrix and $B$ is the observation probability distribution matrix or the emission matrix. Common notations are used in this paper as follows.

a) $S$: A set of $N$ hidden states is denoted as $S = \{s_1, s_2, s_3, \ldots, s_N\}$. The state of model at time $t$ is denoted by $q_t \in S$, which denotes the current state.

b) $V$: A set of $M$ observation states is denoted as $V = \{v_1, v_2, v_3, \ldots, v_M\}$. The observed state at time $t$ is denoted by $o_t \in V$.

c) $\pi$: an $N \times 1$ initial probability distribution over the state. $\pi_i$ is the probability that the Markov chain will start in the state $s_i$. Some states $s_j$ may have $\pi_j = 0$, meaning that they cannot be the initial state. Also, $\sum_{i=1}^{N} \pi_i = 1$.

$$\pi = \{\pi_1, \pi_2, \ldots, \pi_N\}.$$ (4)

d) $A$: $N \times N$ state transition probability matrix, $N$ is the number of hidden states. $A = [a_{ij}], 1 \leq i, j \leq N$.

$$a_{ij} = P(q_{t+1} = s_j|q_t = s_i),$$

(5)

where $a_{ij}$ is the probability of transmission from structural damage state $s_i$ at time $t$ to structural damage state $s_j$ at time $t + 1$ and $\sum_{j=1}^{N} a_{ij} = 1$.

e) $B$: observation probability distribution matrix,

$$b_{jk} = P(o_t = v_k|q_t = s_j),$$

(6)

where $1 \leq j \leq N, 1 \leq k \leq m$, $b_{jk}$ is the possibility of observation $o_t$ at time $t$.

There are three basic problems in HMM: evaluation, decoding, and learning. Given a training set of observations, HMM is trained to find the optimized parameters $\lambda = \{\pi, A, B\}$, by Baum-Welch expectation-maximization algorithm (EM algorithm), which is a learning problem of HMM. After the HMM trained, this model can be applied to evaluate the damage states of a new defective pipeline in various conditional environments.
The BaumWelch algorithm is employed in HMM to perform the training as shown in Fig. 1, which is an iterative process to adjust the parameters of $A$, $B$, and $\pi$.

![Baum-Welch algorithm (EM algorithm)](image)

**Figure 1:** Baum-Welch algorithm (EM algorithm)

### 2.3 Gaussian mixture model

In this paper, the Gaussian mixture model $^{[22, 23]}$, based on unsupervised learning, is chosen to model the distribution of the observed damage indexes to obtain $B$ in the hidden Markov model, whose elements are given by equation

$$b_{jk} = P(o_t = v_k|q_t = S_j) = \sum_{k=1}^{M} c_{jk} N(O_t, \mu_{jk}, \Sigma_{jj}), \quad (7)$$

where $c_{jk}$, $\mu_{jk}$ and $\Sigma_{jj}$ are the mixture weight, mean vector, the covariance matrix of the K-component Gaussian mixture model.

Based on the Gaussian distribution:

$$N(O_t, \mu_{jk}, \Sigma_{jj}) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \quad (8)$$

The EM algorithm is implied in the GMM model, which contains initialization, E step, and M step.
a) Initialization
Setting the number of mixture component $K$, for each component initialize it with $c_{jk}$, $\mu_{jk}$ and $\Sigma_{jz}$.

b) E step
Calculate the posterior probability based on current $c_{jk}$, $\mu_{jk}$ and $\Sigma_{jz}$.

$$\gamma(z_{tk}) = \frac{c_{jk}N(O_t, \mu_{jk}, \Sigma_{jz})}{\sum_{k=1}^{M} c_{jk}N(O_t, \mu_{jk}, \Sigma_{jz})}. \tag{9}$$

$$\mu_{jk} = \frac{1}{N_k} \sum_{t=1}^{N} \gamma(z_{tk})O_t. \tag{10}$$

$$\Sigma_{jz} = \frac{1}{N_k} \sum_{t=1}^{N} \gamma(z_{tk})(O_t - \mu_{jk})(O_t - \mu_{jk})^T. \tag{11}$$

where $N_k = \sum_{t=1}^{N} \gamma(z_{tk})$.

c) M step
Calculate new $c_{jk}$, $\mu_{jk}$, $\Sigma_{jz}$ based on the $\gamma(z_{tk})$.

d) Iteration these steps until the GMM model converges.

### 2.4 GMM-HMM method

There are five major steps included in the damage detection process.

1) States of the HMM are designed based on the placement of pipelines and sensors.

2) Stress waves are obtained from the sensors based on different states of the HMM. Damage indexes are extracted from the stress wave and describe the various degree of damages. The GMM model calculates the parameter of each Gaussian model which formed matrix $B$ in HMM.

3) A typical HMM model defined by a three-tuple: $\lambda = \{\pi, A, B\}$. To initialize the HMM, $\pi$ and $A$ is estimate by the prior probability.

4) After the parameter initialization of the HMM, the training implemented by an iterative process through the Baum-Welch algorithm.

5) Given the trained HMM and the new observations to find out the optimal state is the typical decoding problem in HMM.
3 Experimental Results

3.1 Leakage detection

3.1.1 Experiment setup

The purpose of the experiment is to detect pipeline leakage utilizing PZT sensors. The negative pressure wave is generated by leakage in the pipeline and propagates along the pipeline from the leakage point to both ends.

The experimental pipeline built at the University of Houston, as shown in Fig. 2 consists of a series of PVC plain-end pipe sections connected together to form a pipeline with a total length of 55.78 m. Six PZT sensors (P1 to P6, size is 15 mm × 10 mm) are directly mounted on the pipeline to detect NPW signal arrival. A NI PXI-5105 digitizer is used as a data acquisition system. The digitizer is triggered by the voltage signal of PZT No. 1 with the trigger level at −.02 V and all the signals from six PZT sensors are recorded simultaneously at a sampling rate of 100 KS/s. Although it is not targeted to solve the problem of submarine pipeline leakage detection, the experimental data is used to validate the proposed GMM-HMM method.

Two states, leaking or not leaking, are chosen as the states in a left to right HMM model, where the transitions only go from one state to itself or to a unique follower. Leakage signal collected by P1 and baseline simulation signal of three different sections of leakage serves as original data of the model.
3.1.2 Results of leakage detection

Signals of the different sections are shown in Fig. 3(a). The blue dashed line represents leakage state signal and red solid line represents no leakage state signal. The experiment of collecting signals for each leakage state was repeated by 20 times. Randomly selected repeated experiment data from the same sensor are separated into two groups equally, named training data and test data.

![Different states signal](image1.png)

(a) Different states signal

![Damage indexes](image2.png)

(b) Damage indexes

![Negative log-likelihood](image3.png)

(c) Negative log-likelihood

![Pipeline status output](image4.png)

(d) Pipeline status output

Figure 3: Experimental results of leakage detection

For training data, a total of 70 groups, 100000 per group samples are collected. Two damage indexes in different damage leakage states are obtained by using Eq. 1 and Eq. 2 with 45 data points for each state as shown in Fig. 3(b). The parameter of GMM can be calculated by EM algorithm. The negative log-likelihood predicted by GMM shown in Fig. 3(c) The testing is performed after the training process of the HMM model whose parameters have been optimized. The testing is used to validate the prediction capability of the HMM model and the result is showed in Fig. 3(d). The accuracy of the testing performance between predicted state sequence and
actual state sequence are approaching 92.51%. The accuracy is the ratio of correctly predicted states to the total number of state sequence.

3.2 Leakage location detection

3.2.1 Experiment setup

In this experiment, all the data are based on the previous experimental setup. By changing the leakage state to different leakage locations. The purpose and output of the HMM model will be changed to leakage location detection. As shown in Fig. 4(a), there are three leakage locations corresponding to three different states in ergodic HMM. State 1 denotes leakage occurred at section 1 of pipeline, state 2 and 3 denotes leakage occurred at section 2 and 3 of pipeline. Therefore, there are three states in this ergodic HMM, which allowing for transitions from any emitting state to any other emitting state.

![Schematic diagram of pipeline statuses](image1)

(a) Schematic diagram of pipeline statuses

![Signal of different states](image2)

(b) Signal of different states

Figure 4: Experimental results of leakage location detection

Leakage signals are collected by $P_1$. As shown in Fig. 4(b), a total of 100 groups, 200000 per group samples are collected. To reduce the computational
volume, data for this experience are cropped from the original data by 90000 data points to each state.

### 3.2.2 Results of leakage location detection

The 100 groups of measurements from three different leakage locations are separated into a training group and a test group. According to the observation density estimation method, the parameters of the Gaussian mixture model for each leakage state can be calculated by GMM based on EM algorithm. The maximum-likelihood number of Gaussian component was 3 estimated by GMM. The parameters of the HMM model are initialized as:

$$\pi_0 = [1, 0, 0], A = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$  

The parameters of the three states HMM model are reassessed by the Baum-Welch algorithm. After training the updated parameters obtained are as follows:

$$\pi = [0, 0, 1], A = \begin{bmatrix} 0.232 & 0.357 & 0.411 \\ 0 & 0.525 & 0.475 \\ 0.539 & 0.102 & 0.359 \end{bmatrix}.$$  

Then, the testing is performed after the parameter optimization of HMM model. The model calculates the posterior probability when new testing data are fed. By maximizing the posterior probability, the predicted state will be obtained. The leakage location evaluation result is shown in Fig. 5. The accuracy of the testing performance between predicted state sequence and actual state sequence are approaching 94.81%.

### 3.3 Crack depth inspection

#### 3.3.1 Experiment setup

The purpose of the experiment is to detect the depth of cracks utilizing PZT sensors. The experimental pipeline built at the University of Houston, consists of a section of galvanized steel pipe with a total length of 3 meters. A PZT array with sixteen PZT transducers are directly mounted on the pipeline to detect deflection signal. Shown in Fig. 6. The guided wave is generated.
from sensors at the left side of pipe at the frequency of 50 kHz, propagates along the pipeline through the defect point, and received by the sensors at other side. This PZT array can detect the pipeline defect location based on time-reversal method and matching pursuit de-noising [24]. In addition, this PZT array has been used for underwater communication by using stress wave propagation along pipelines [25].

Three states, defect depth 1, defect depth 2, and defect depth 3, are chosen as the states in the left to right HMM model, which means the depth of the defect was increased unidirectionally. Received defect signal is collected from the right-hand side sensors.

### 3.3.2 Results of depth inspection

The 154 group measurements from three defect depth inspection are separated into a training group and a test group. The parameters of the Gaussian mixture model are also calculated based on the observation density estimation method, where the number of Gaussian is set to 3. The parameters of the left to right HMM model are initialized as:

![Figure 5: Actual state vs. HMM predict state](image)
The parameters of the three states HMM model are also reassessed by the Baum-Welch algorithm. After training the updated parameters obtained are as follows:

\[ \pi_0 = [1, 0, 0], \quad A = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.9 & 0.1 \\ 0 & 0 & 1 \end{bmatrix}. \]

By maximizing the posterior probability, the predicted state will be obtained. The crack depth inspection result is shown in Fig. 7. The accuracy of testing performance between predicted state sequence and actual state sequence are approaching 93.23%.

### 4 Conclusions

Pipeline leakage detection and crack depth identification are difficult especially in changing environment and time-varying operational conditions. This
study applied a GMM-HMM based on a probabilistic damage evaluation method on pipeline leakage detection. The GMM-HMM method calculates the maximum log-likelihood recursively, which reveals the emission probability from hidden states to observable states. Thus, HMM can have a better description of the leakage by using observations from GMM which contains implicit damage indexes. The experiment performed on the pipeline showed great potential to deal with various leakage detection problems.

**Funding**  This research was funded by National Science Foundation grant number 1801811.

**Acknowledgments**  Thanks to Smart Materials and Structures Laboratory at the University of Houston for providing the pipeline leakage experimental data.

**References**

[1] M. Ho, S. El-Borgi, D. Patil, and G. Song, “Inspection and monitoring systems subsea pipelines: A review paper,” *Structural Health Monitoring*, vol. 19, no. 2, pp. 606–645, 2020.
[2] M. A. Adegboye, W.-K. Fung, and A. Karnik, “Recent advances in pipeline monitoring and oil leakage detection technologies: principles and approaches,” Sensors, vol. 19, no. 11, p. 2548, 2019.

[3] G. Zhang, S. C. M. Ho, L. Huo, and J. Zhu, “Negative pressure waves based high resolution leakage localization method using piezoceramic transducers and multiple temporal convolutions,” Sensors, vol. 19, no. 9, p. 1990, 2019.

[4] J. Zhu, L. Ren, S.-C. Ho, Z. Jia, and G. Song, “Gas pipeline leakage detection based on PZT sensors,” Smart Materials and Structures, vol. 26, no. 2, p. 025022, 2017.

[5] Q. Hou, L. Ren, W. Jiao, P. Zou, and G. Song, “An improved negative pressure wave method for natural gas pipeline leak location using FBG based strain sensor and wavelet transform,” Mathematical Problems in Engineering, vol. 2013, pp. 1–8, 2013.

[6] W. Liang, L. Zhang, Q. Xu, and C. Yan, “Gas pipeline leakage detection based on acoustic technology,” Engineering Failure Analysis, vol. 31, pp. 1–7, 2013.

[7] Y. Li and L. Sun, “Leakage detection and location for long range oil pipeline using negative pressure wave technique,” in 2009 4th IEEE Conference on Industrial Electronics and Applications. IEEE, may 2009.

[8] Wikipedia contributors, “Leak detection — Wikipedia, the free encyclopedia,” 2020, [Online; accessed 13-August-2020]. [Online]. Available: https://en.wikipedia.org/wiki/Leak_detection

[9] N. Li, X. Chen, and G. Song, “Fatigue crack width detection based on the active sensing method: A feasibility study,” in 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), IEEE. IEEE, may 2019, pp. 186–189.

[10] C. Wang, N. Wang, S.-C. Ho, X. Chen, M. Pan, and G. Song, “Design of a novel wearable sensor device for real-time bolted joints health monitoring,” IEEE Internet of Things Journal, vol. 5, no. 6, pp. 5307–5316, 2018.
[11] P. Gong, M. Luo, L. Zhou, L. Jiang, and X. Chen, “An image processing method for extraction of the stress wave reflection period,” *Applied Sciences*, vol. 10, no. 10, p. 3486, may 2020.

[12] D. Chakraborty, N. Kovvali, A. Papandreou-Suppappola, and A. Chattopadhyay, “An adaptive learning damage estimation method for structural health monitoring,” *Journal of Intelligent Material Systems and Structures*, vol. 26, no. 2, pp. 125–143, 2015.

[13] J.-H. Yang and H.-F. Lam, “An efficient adaptive sequential Monte Carlo method for Bayesian model updating and damage detection,” *Structural Control and Health Monitoring*, vol. 25, no. 12, p. e2260, 2018.

[14] Z. Li, J. Han, J. Sun, Y. He, and F. Chu, “Fault recognition method based on independent component analysis and hidden markov model,” *Journal of Vibration and Control*, vol. 13, no. 2, pp. 125–137, feb 2007.

[15] L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[16] A. D. Wilson and A. F. Bobick, “Parametric hidden markov models for gesture recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 884–900, 1999.

[17] M. A. Mohamed and P. Gader, “Generalized hidden Markov models. II. Application to handwritten word recognition,” *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 1, pp. 82–94, 2000.

[18] T. Xia and X. Chen, “A discrete hidden markov model for SMS spam detection,” *Applied Sciences*, vol. 10, no. 14, p. 5011, jul 2020.

[19] R. Rammohan and M. R. Taha, “Exploratory investigations for intelligent damage prognosis using hidden Markov models,” in *2005 IEEE International Conference on Systems, Man and Cybernetics*, vol. 2. IEEE, 2005, pp. 1524–1529.

[20] C. Tschope and M. Wolff, “Statistical classifiers for structural health monitoring,” *IEEE Sensors Journal*, vol. 9, no. 11, pp. 1567–1576, 2009.
[21] S. Torkamani, S. Roy, M. E. Barkey, E. Sazonov, S. Burkett, and S. Kotru, “A novel damage index for damage identification using guided waves with application in laminated composites,” *Smart Materials and Structures*, vol. 23, no. 9, p. 095015, 2014.

[22] H. Mei, S. Yuan, L. Qiu, and J. Zhang, “Damage evaluation by a guided wave-hidden Markov model based method,” *Smart Materials and Structures*, vol. 25, no. 2, p. 025021, 2016.

[23] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[24] Y. Xu, M. Luo, Q. Liu, G. Du, and G. Song, “PZT transducer array enabled pipeline defect locating based on time-reversal method and matching pursuit de-noising,” *Smart Materials and Structures*, vol. 28, no. 7, p. 075019, jun 2019.

[25] S. He, N. Wang, M. Ho, J. Zhu, and G. Song, “Design of a new stress wave communication method for underwater communication,” *IEEE Transactions on Industrial Electronics*, 2020.