Weakly Supervised Defect Detection using Acoustic Data based on Positive and Negative Constraints

Jun Younes Louhi Kasahara[a], Atsushi Yamashita[a], Hajime Asama[a]

[a]Department of Precision Engineering, The University of Tokyo, Japan
E-mail: louhi@robot.t.u-tokyo.ac.jp

Abstract -

Concrete structures are heavily used in most modern societies and the population of structures in need of inspection is rapidly growing. On the other hand, the manpower for inspection is decreasing. This has brought into focus the need for automated inspection methods for concrete structures. The hammering test is a popular method for inspection that uses the sound resulting from a hammer impact on the surface of the structure for defect detection. Previous methods largely employed machine learning approaches for the automation of the hammering test. Weakly supervised methods used positive queries answers on sample pair similarity: a human user was questioned on the similarity of pairs of hammering samples and similar pairs were used to transform the feature space. However, it can be expected that dissimilar pairs would also be gathered in this process. Therefore, in the present paper is proposed a method for weakly supervised defect detection in concrete structures using hammering with both positive and negative answers to queries. After the initial feature space transformation based on positive query answers, another feature space transformation is introduced based on negative query answers. Experiments in laboratory conditions showed the effectiveness of the proposed method.

Keywords -

Defect detection; Infrastructure inspection; Weak supervision; Acoustic data; Clustering

1 Introduction

Concrete structures are featured heavily in most modern societies. This is especially true for social infrastructures such as tunnels, highways and bridges. Due to various factors ranging from simple aging to damage caused by environmental conditions, concrete structures require regular and careful inspection. This is a critical aspect for social infrastructures due to their large number of users, for which safety is of utmost concern [1]. Recent events such as the collapse of the Sasago tunnel in Japan [2] or the collapse of the Morandi bridge in Italy [3] have emphasized the issues caused by an ever-increasing population of aging structures facing an ever-decreasing population of workers tasked with inspection work.

Figure 1. A human inspector conducting the hammering test on the wall portion inside a tunnel. The hammer is used to hit the surface of the concrete structure and the returned sound used to assess the presence of defect beneath the surface.

One popular inspection method for concrete structures is the hammering test, illustrated in Figure 1. It consists in a human inspector using a simple hammer to hit the surface of the structure and assessing the presence of defects beneath the surface from the impact sound. Due to both its simplicity and non-destructive nature, it is widely popular. However, it requires a skilled human inspector to be able to distinguish impact sounds resulting from defect portions of the structures. Due to the manpower shortage and the population of structures in need of testing, the hammering test in its current, traditional form, is not effective. Therefore, the automation of the hammering test is highly sought after and has attracted much focus in recent years [4][5][6].

Previous works have mainly employed machine learning approaches to tackle this issue. Supervised learning approaches use training data to train classifiers to distinguish defect and non-defect sounds. In [7], a Neural Network was used with a Radial Basis Function in order to detect defects in concrete bridges by the sound of dragging chains. In [8], Time-Frequency Analysis was employed together with Ensemble Learning and achieved accurate classification of hammering samples into defect and non-defect classes. Furthermore, classification of defects samples into shallow and deep classes was also achieved. Supervised learning approaches often boasted remarkable performance. However, they are limited by the availability of adequate training data. Indeed, if the training data does not correspond to the actual tested concrete structure, difficulties arises when attempting to produce a good model. Since concrete is an aggregate, each structure is unique and therefore, besides the fact that generating training data is a human labor intensive task, adequate training data may only be obtainable on site, i.e., on the very tested structure. This severely limits the practicability of such approaches.

Unsupervised learning methods, characterized by the fact that they to not require training data, offer an interesting alternative to this issue. In [9], clustering was used on Fourier spectrum of hammering samples with a correlation distance. In [10] and [11], clustering of audio hammering samples was considered along each hammering sample's hit location using a camera. Unsupervised learning approaches successfully bypassed the practical issue caused by training data. However, they generally have lacking performance compared to supervised learning methods in their optimal conditions. Furthermore, Unsupervised Learning approaches also incorporate strong priors, which requires careful consideration in the design phase.

Between supervised and unsupervised learning methods, weakly supervised approaches aims to combine the best of both worlds by only requiring *weak supervision*, i.e., a form of supervision which is less informative but also presents less burden on the human user than generating training data. In [12], an initial framework for the automation of the hammering test based on pairs of hammering samples a human user has indicated as similar was proposed. In [13], this initial framework was reinforced by the addition of hammering samples' hit location using a camera. In [14], an active query scheme was proposed to ensure more consistent weak supervision quality. Good results were obtained. However, only positive answers to queries were considered in those works: weak supervision is gathered from human users, through queries on sample pair similarity. Positive answers to such queries, i.e., the sample pair is similar, are known as *must-links*. Negative answers, i.e., the sample pair is dissimilar, are known as *cannot-links*. It is realistic to assume that the human user is limited in the number of queries it can answer to and while approaches such as [14] attempted to maximize the number of obtained must-links through active query, all the queries resulting in must-links cannot be ensured. This means that there will inevitably be cannot-links generated during the query process, which are not taken advantage of in previous work regarding automation of hammering test.

Therefore, in this paper, the objective is to achieve defect detection in concrete structures using acoustic data with both positive and negative answers to queries.

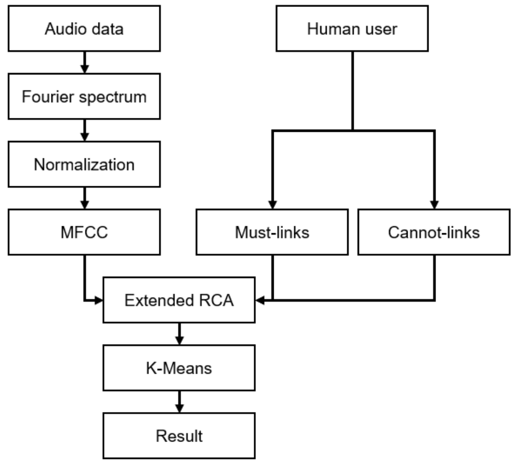


Figure 2. Overview of the proposed method.

2 Method

2.1 Overview

An overview of the proposed method is shown in Figure 2. The input, audio data, is first pre-processed: this involves a conversion to Fourier spectrum, normalization and conversion to Mel-Frequency Cepstrum Coefficients (MFCC). Then, weak supervision, provided by a human user under the form of must-links and cannot-links, is used to conduct a transformation of the feature space to match the human user's notion of similarity. Finally, separation between defect and non-defect samples is conducted by clustering using K-Means.

2.2 Initial Feature Space

Hammering samples are initially collected as audio segments. The first step of the processing is conversion to Fourier spectrum. Given a sound sample defined by $\mathbf{x} = (x_1, ..., x_d)$, its Fourier spectrum $\mathbf{a} = (a_1, ..., a_d)$ is calculated. Next, since there is no assumption about regularity of the input, i.e., the hammer strike is not assumed to be of constant force, a normalization to zero mean and unit variance is conducted as in (2), with $\bar{a}$ being the mean of the components of a_i as defined in (1).

$$\bar{a} = \frac{1}{d}\sum_{i=1}^{d} a_i, \tag{1}$$

$$a_i = \frac{a_i - \bar{a}}{\sqrt{\frac{\sum_{i=1}^{d}(a_i - \bar{a})^2}{d-2}}}. \tag{2}$$

In [10], the MFCC feature space was shown to be a good feature space for discrimination of defect hammering samples. MFCC are hand-crafted feature vectors originally build for speech recognition and popular across several fields dealing with audio data [15, 16]. MFCC are suited for hammering samples because they emulate the human hearing and the human hearing is able to discriminate defect hammering samples from non-defect hammering samples.

To compute the MFCC, first the periodogram estimate of the power spectrum is computed. Then, what are known as Mel filterbanks, a set of N_{filter} triangular filters equally spaced in the Mel scale, are applied. The Mel scale is an empirical scale tuned to the sensitivity of the human cochlea, as in (3) with f the frequency in Hertz. The logarithm of the resulting N_{filter} energy values are further processed by Discrete Cosine Transform to finally obtain MFCC.

$$M(f) = 1125 * \ln(1 + \frac{f}{700}). \quad (3)$$

For the sake of clarity, in the remainder of this paper, the MFCC feature vector of a hammering sample will simply be noted as $\mathbf{x}$.

2.3 Weakly Supervised Feature Space Transformation: Extended Relevant Component Analysis

The feature space defined by MFCC is a good one for separating defect and non-defect hammering samples. However, improvements can be achieved by further transforming the feature space to match the human user's notion of similarity according to answers provided to queries. Those answers to queries can come in two forms: pairs of samples the human user considers similar are called must-links whereas pairs of samples the human user considers dissimilar are called cannot-links.

The previous work [14] only employed must-links through Relevant Component Analysis (RCA). RCA is a weakly supervised metric learning method initially proposed in [17]. While the authors in [17] puts the fact that RCA is only based on must-links as an advantage, as must-links are easier to generate compared to cannot-links, in practice it can be reasonably expected that the human user answering queries would be limited in the number of queries he can answer to, rather than in the number of must-links he can provide. This means that the querying process is very likely to produce cannot-links along must-links. Therefore, not using cannot-links to contribute to the feature space transformation is wasteful.

Extended RCA is an extension of the original RCA initially proposed in [18]. The feature space transformation matrix build upon must-links $\mathcal{M}$ differs slightly from what is used in RCA, as shown in (4). This allows to build a similar transformation matrix on the set of cannot-links C as well, as in (5).

$$\hat{\mathbf{C}}_{\mathcal{M}} = \frac{1}{2|\mathcal{M}|} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{M}} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \quad (4)$$

$$\hat{\mathbf{C}}_{C} = \frac{1}{2|C|} \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in C} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \quad (5)$$

$$\mathbf{y}_i = \hat{\mathbf{C}}_{C}^{1/2}\hat{\mathbf{C}}_{\mathcal{M}}^{-1/2}\mathbf{x}_i, \quad 1 \le i \le N. \quad (6)$$

Therefore, each sample $\mathbf{x}_i$ is first linearly transformed to $\mathbf{y}_i^{'} = \mathbf{C}_{\mathcal{M}}^{-1/2}\mathbf{x}_i$, using must-links, and then linearly transformed again to $\mathbf{y}_i = \mathbf{C}_{C}^{1/2}\mathbf{y}_i^{'}$, using cannot-links this time. The transformation based on $\hat{\mathbf{C}}_{\mathcal{M}}$ aims to reduce the within-class scatter while the transformation based on $\hat{\mathbf{C}}_{C}$ aims to increase the between-class scatter.

2.4 Clustering

After the weakly supervised feature space transformation described in the previous section, separation of hammering samples between defect and non-defects is conducted using K-Means [19]. K-Means is simple iterative clustering algorithm that aims to achieve a partitioning of the dataset by minimizing the variance of each cluster S_k. Algorithm 1 shows a pseudo-algorithm of K-Means.

3 Experiments

Experiments were conducted in laboratory conditions using concrete test blocks containing simulated defects. The used setup is illustrated in Figure 3. The blocks were hit on several locations, once per location, using a KTC UDHT-2 hammer (head diameter 16 mm, length 380 mm, weight 160 g). This hammer is commonly used in actual inspection sites. Audio recording was done using a Behringer ECM8000 microphone fixed roughly at 0.5 m from the concrete test block and coupled with a Roland

Algorithm 1: Pseudo-algorithm of K-Means.

Data: Dataset D of N samples $\mathbf{y}_i$,
number of clusters K
Result: Partition of D into K clusters
Initialization:
Initialize cluster centroids $\mathbf{c}_1$,$\mathbf{c}_2$,...,$\mathbf{c}_K$ randomly;
while termination criterion not met **do**
Assign samples:
for each sample $\mathbf{y}_i$
$l_i \rightarrow \underset{k}{\mathrm{argmin}}||\mathbf{y}_i - \mathbf{c}_k||^2$
Update centroids:
for each centroid $\mathbf{c}_k$
$\mathbf{c}_k \rightarrow \frac{1}{|S_k|}\sum_{\mathbf{y}_i\in S_k}\mathbf{y}_i$
end

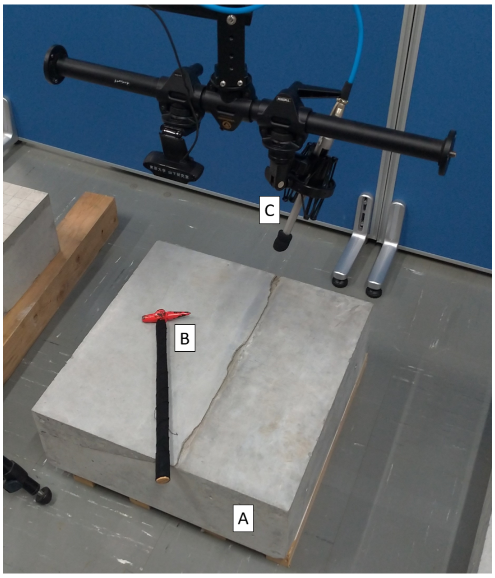
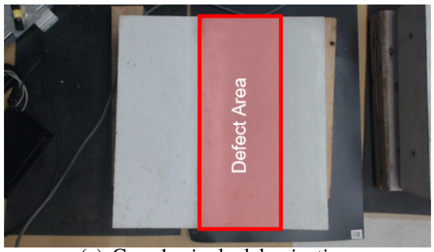
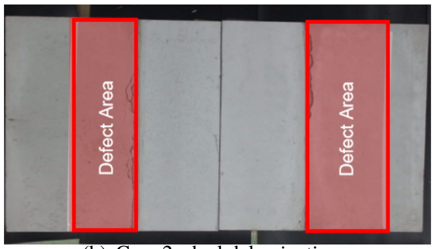


Figure 3. The experimental setup in laboratory conditions with (A) a concrete test block containing a simulated defect, (B) a hammer and (C) a microphone to record the hammering sound.



(a) Case 1: single delamination.



(b) Case 2: dual delaminations.

Figure 4. Concrete test blocks used in laboratory experiments containing man-made defects. Defect areas are therefore precisely known and indicated by light red overlays.

UA-25EX sound board at 44.1 kHz. MFCC were computed with 10 coefficients. 30 queries were allowed for each run.

Two scenarios were considered, those are the same as featured in [14] and the setting $K = 2$ was used for both:

- Case 1: single delamination. This dataset contains 462 samples: 272 non-defects and 190 defects. The delamination is at an angle of 30 degrees.
- Case 2: dual delaminations. This dataset contains 270 samples: 155 non-defects and 115 defects. Two delaminations are present, both at an angle of 15 degrees.

A picture for both cases is provided in Figure 4.

4 Results and Discussion

In Table 1 is reported the average number of must-links and cannot-links obtained out of 30 queries over 50 runs. About half of the queries effectively resulted in cannot-links, which are essentially wasted queries for previous approaches. Since sample pairs were queried randomly, the ratio of must-links and cannot-links reflected approximately the datasets' ratio of defect/defect, non-defect/non-defect pairs and defect/non-defect pairs.

In Figure 5 are reported the average performance obtained over 50 runs in both considered cases for the approach of [12] and the proposed method. Error bars correspond to one standard deviation. The performance was measured using the Rand index [20]. The Rand index is a common measure of performance for clustering methods and is essentially a ratio of correctly clustered sample pairs over the total number of sample pairs in the dataset. It ranges between 0 and 1, with higher values of Rand index indicating the better clustering.

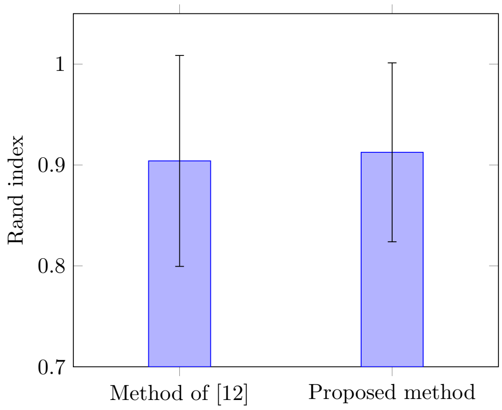
For both Case 1 and Case 2, it can be noticed that the proposed method achieved better clustering on average than the method of [12]. The spread of the output seems to be also narrower for the proposed method, indicated by lower values of standard deviation. This is especially noticeable for Case 2. This is certainly due to the increased number of constraints used in the feature space transformation by the proposed method, defining more precisely the target feature space.

The performance of the method of [12] on Case 1 is significantly lower than reported in the initial publication. This is due to the difference in number of effective must-links: while 20 must-links were allowed in [12] whereas in the present paper 30 queries were allowed, resulting in about only 15 must-links in average. This indicates that, depending on the dataset, a significantly larger number of queries is potentially required to obtain the desired number of must-links.

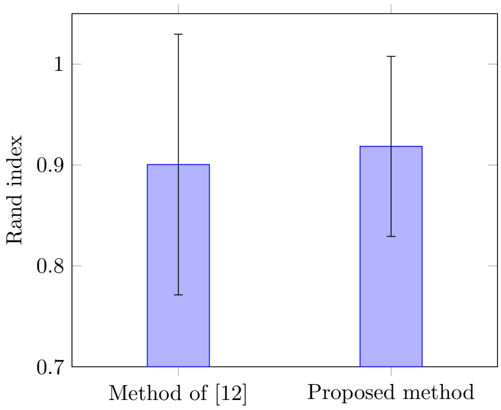
With the same number of must-links, the proposed method, that makes use of cannot-links as well, does perform better. However, the performance increase enabled by the additional feature space transformation computed based on cannot-links does not bring as much improvement

Table 1. Average number of constraints of each type over 50 runs for 30 allowed queries.

	Number of allowed queries	**Average number of obtained must-links**	**Average number of obtained cannot-links**
Case 1: single delamination	30	15.01	14.99
Case 2: dual delaminations	30	15.41	14.59



(a) Case 1: single delamination.



(b) Case 2: dual delaminations.

Figure 5. Performance of the method of [12] and the proposed method. Average over 50 iterations reported, error bars correspond to one standard deviation. 30 queries were allowed and random seeding was used for each run.

per constraint compared to must-links. This potentially indicates that must-links are more informative and better suited for fine-tuning the feature space, at least within the RCA framework.

5 Conclusion

In the present paper was proposed a method for defect detection in concrete structures using acoustic data based on both positive and negative constraints. Using Extended RCA, an additional feature space transformation based on negative constraints was conducted following the first feature space transformation using positive constraints. Experiments in laboratory conditions using concrete test blocks showed that the proposed method achieved better results more consistently than the previous method that only used positive constraints.

As future work, we would like to further study the influence of negative constraints on the final feature space. As unavoidable by-products of the querying process, cannot-links should be employed to maximize the data provided by the human user. However, a straight inclusion of those negative constraints in the RCA framework might not be their only use. For example, negative constraints obtained early in the query process could be used in the selection process for the next sample pair to query the human user on.

Acknowledgments

This work was supported in part by the Japan Construction Information Center (JACIC) foundation. The authors thank Prof. Hiromitsu Fujii, Department of Advanced Robotics, Chiba Institute of Technology for his valuable help. The authors also thank the Institute of Technology, Tokyu Construction Co., Ltd. for their precious help and cooperation.

References

[1] Keiji Nagatani and Yozo Fujino. Research and development on robotic technologies for infrastructure maintenance. In Journal of Robotics and Mechatronics, volume 31, pages 744–751. Fuji Technology Press Ltd., 2019.

[2] BBC News. Japan Sasago tunnel collapse killed nine. [web page]. https://www.bbc.com/news/world-asia-20576492, 2012. Accessed: 2019-04-11.

[3] BBC News. Italy bridge collapse: Genoa death toll rises to 43. [web page]. `https://www.bbc.com/news/world-europe-45241842`, 2018. Accessed: 2019-04-11.

[4] Yusuke Takahashi, Satoru Nakamura, Yasuichi Ogawa, and Tomoya Satoh. Velocity control

mechanism of the under-actuated hammering robot for gravity compensation. In Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC), 2017.

[5] Satoru Nakamura, Yusuke Takahashi, Daisuke Inoue, and Takao Ueno. The variable guide frame vehicle for tunnel inspection. In Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC), 2017.

[6] Yusuke Takahashi, Satoshi Maehara, Yasuichi Ogawa, and Tomoya Satoh. Concrete inspection systems using hammering robot imitating sounds of workers. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, volume 35, pages 1–5. IAARC Publications, 2018.

[7] Gang Zhang, Ronald S. Harichandran, and Pradeep Ramuhalli. An automatic impact-based delamination detection system for concrete bridge decks. In NDT & E International, volume 45, pages 120–127. Elsevier, 2012.

[8] Hiromitsu Fujii, Atsushi Yamashita, and Hajime Asama. Defect detection with estimation of material condition using ensemble learning for hammering test. In Proceedings of the International Conference on Robotics and Automation (ICRA), pages 3847–3854. IEEE, 2016.

[9] Jun Younes Louhi Kasahara, Hiromitsu Fujii, Atsushi Yamashita, and Hajime Asama. Unsupervised learning approach to automation of hammering test using topological information. In ROBOMECH Journal, volume 4, pages 13–23. Springer, 2017.

[10] Jun Younes Louhi Kasahara, Hiromitsu Fujii, Atsushi Yamashita, and Hajime Asama. Clustering of spatially relevant audio data using Mel-frequency cepstrum for diagnosis of concrete structure by hammering test. In Proceedings of the International Symposium on System Integration, pages 787–792. IEEE, 2017.

[11] Jun Younes Louhi Kasahara, Hiromitsu Fujii, Atsushi Yamashita, and Hajime Asama. Fuzzy clustering of spatially relevant acoustic data for defect detection. In Robotics and Automation Letters, volume 3, pages 2616–2623. IEEE, 2018.

[12] Jun Younes Louhi Kasahara, Hiromitsu Fujii, Atsushi Yamashita, and Hajime Asama. Weakly supervised approach to defect detection in concrete structures using hammering test. In Proceedings of the International Conference on Consumer Electronics, pages 997–998, 2019.

[13] Jun Younes Louhi Kasahara, Hiromitsu Fujii, Atsushi Yamashita, and Hajime Asama. Complementarity of sensors and weak supervision for defect detection in concrete structures. In Proceedings of the International Symposium on System Integration, pages 1–6. IEEE, 2020.

[14] Jun Younes Louhi Kasahara, Atsushi Yamashita, and Hajime Asama. Acoustic inspection of concrete structures using active weak supervision and visual information. In Sensors, volume 20, page 629. Multidisciplinary Digital Publishing Institute, 2020.

[15] Todor Ganchev, Nikos Fakotakis, and George Kokkinakis. Comparative evaluation of various MFCC implementations on the speaker verification task. In Proceedings of the International Conference on Speech and Computer (SPECOM), volume 1, pages 191–194. International Speech Communication Association, 2005.

[16] Meinard Müller. Information retrieval for music and motion. Springer, 2007.

[17] Aharon Bar-Hillel, Tomer Hertz, Noam Shental, and Daphna Weinshall. Learning distance functions using equivalence relations. In Proceedings of the International Conference on Machine Learning (ICML), pages 11–18, 2003.

[18] Dit-Yan Yeung and Hong Chang. Extending the relevant component analysis algorithm for metric learning using both positive and negative equivalence constraints. In Pattern Recognition, volume 39, pages 1007–1010. Elsevier, 2006.

[19] Stuart Lloyd. Least squares quantization in PCM. In Transactions on Information Theory, volume 28, pages 129–137. IEEE, 1982.

[20] William M Rand. Objective criteria for the evaluation of clustering methods. In Journal of the American Statistical Association, volume 66, pages 846–850. Taylor & Francis Group, 1971.