Filtering Queries for Weakly Supervised Clustering Using Visual Information for Defect Detection in Concrete Structures

○ ルイ笠原純ユネス†，山下淳†，淺間一†

○ Jun Younes LOUHI KASAHARA†, Atsushi YAMASHITA†, Hajime ASAMA†

†：東京大学，louhi@robot.t.u-tokyo.ac.jp

＜要約＞　Automatic inspection of concrete structures using acoustic methods such as the hammering test is highly demanded due to the aging of social infrastructures in most modern societies: such structures require regular inspection in order to guarantee their users' safety.　From this aspect, weakly supervised methods are attractive since they also to preserve a subtle mix of human involvement in the inspection process. In this paper is proposed a method to filter queries using visual information for such weakly supervised methods. This allows for more efficient collection of weak supervision and, therefore, allows to increase performance. Experiments conducted using man-made concrete test blocks showed the effectiveness of the proposed method.

＜キーワード＞ Weak supervision,　Clustering,　Defect detection

1.　Introduction

Concrete is a heavily used material in most modern societies, especially in social infrastructures such as tunnels, bridges and highways. Catastrophic events in recent history have underlined the need to regularly inspect such structures in order to ensure their users' safety [1]. One popular method for such task is known as the hammering test.

Previous works aiming at the automation of the hammering test mainly employed machine learning approaches. In [2], another acoustic method consisting in dragging chains was used with Linear Prediction Coefficients. In [3], Neural Networks were used and in [4], Ensemble Learning along time-frequency analysis was considered. Such approaches are known as supervised learning and exhibited strong performance but their practicability was limited due their heavy reliance on training data, which are both laborious and difficult to obtain.

The work in [5][6] used clustering, an unsupervised learning method, which does not have such drawbacks but ultimately suffered from lower performance.

In our previous work, a weakly supervised framework for the automation of the hammering test, where weak supervision was collected by querying a human user on sample pairs, was proposed [7]. However, since all queries would not be positive, the considered settings of a fixed amount of usable weak supervision was not realistic.

Therefore, in the present paper is proposed a filtering method for weak supervision queries using visual information for defect detection in concrete structures in order to increase the efficiency of weak supervision generation and increase performance.

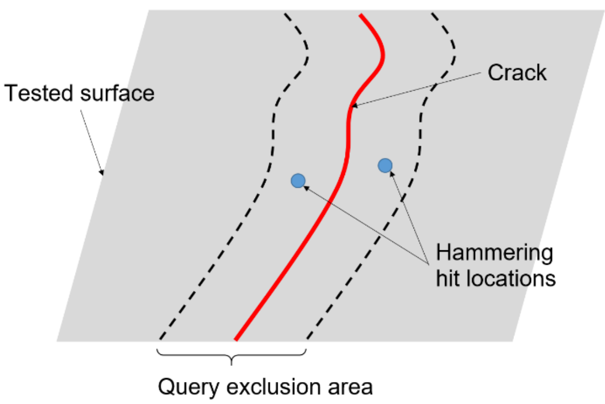

Fig. 1. Illustration of the proposed concept of query exclusion area. By forbidding queries on close samples across cracks, the probability of finding dissimilar samples can be reduced, thus increasing query efficiency.

2. Method

2.1 Query Exclusion Area

Weak supervision is comprised of must-links and cannot-links, which refer to pairs of samples a human has determined to be similar and dissimilar, respectively. For methods such as [7], and more generally weakly supervised methods that does not make use of cannot-links, queries on such pairs are essentially wasted and result in lesser amounts of obtained must-links.

Therefore, the proposed method of this paper is to filter out pairs that are more likely to be cannot-links from potential queries. Most defects appear partially on the surface, as cracks. Cracks usually propagate deeper inside the structure diagonally. This means one side of the crack on the surface has a high probability of being a defect. Therefore, querying on pairs of hammering samples located in close proximity each another and across a crack on the surface has a high probability of resulting in a cannot-link, i.e., wasting a query. The idea of the proposed method is to forbid such queries to allow for increased efficiency, by establishing a query exclusion area, as shown in Fig. 1.

2.2 Audio Pre-Processing

Audio data of hammering samples are converted into Fourier spectrums and normalized to zero mean and unit variance to account for the lack of hammering force monitoring. Then, MFCC feature vectors, used successfully in [6], are computed. An overview of the overall processing flow is shown in Fig. 2.

2.3 Filtering Queries Using Visual Information

First, an image of the inspected area is collected and crack positions are detected. Then, hammering audio samples are collected by hitting several locations and pairs of samples are randomly selected to be considered as potential query.

In the decision process is considered the relative position of sample pairs according to detected cracks: if samples are located on the same side of a crack, they are accepted as queries. If samples are located across a crack, an exclusion zone is defined around the crack so as to eliminate the most potential cannot-links from consideration, as in (1), with $\boldsymbol{x}_1, \boldsymbol{x}_2$ pairs of samples and d_c corresponding to the physical distance to the closest pixel detected as crack:

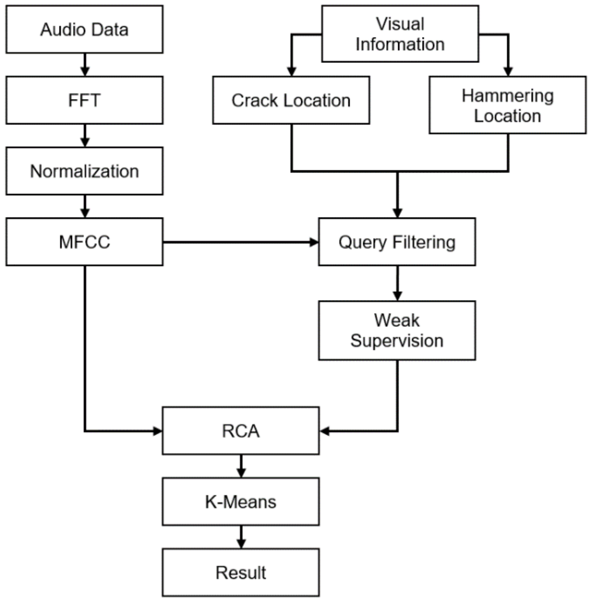

Fig. 2. Overview of the proposed method

$$d_c(\boldsymbol{x}_1) + d_c(\boldsymbol{x}_2) < d_{\min}. \quad (1)$$

This effectively translates to forbidding the consideration of pairs of samples too closely located across a visible crack on the surface of the tested area, with $d_{\min}$ being a parameter controlling the width of this exclusion area. This is practical since inspection requirements usually incorporates a minimum defect size to be detected and this parameter can be set up to reflect this.

2.4 Weakly Supervised Clustering

After the filtering process described earlier, pairs of samples are submitted to the human user as queries to determine if they are must-links or cannot-links. After generation of a set of must-links, the N_{Total} samples involved are used to generate a set of chunklets $\{M_l\}_{l\in[1\ldots N_{chunklet}]}$ by using the transitive property of must-links. Then, with $\widehat{\boldsymbol{m}}_l$ the mean of chunklet M_l, the within chunklet covariance matrix is calculated as in (2). This matrix is used for the whitening transformation as in (3):

$$\widehat{\mathbf{C}} = \frac{1}{N_{Total}} \sum_{l=1}^{N_{chunklet}} \sum_{x_i \in M_l} (\boldsymbol{x}_i - \widehat{\boldsymbol{m}}_l)(\boldsymbol{x}_i - \widehat{\boldsymbol{m}}_l)^T, \quad (2)$$

$$\boldsymbol{x}_i \rightarrow \widehat{\mathbf{C}}^{-1/2} \boldsymbol{x}_i. \quad (3)$$

This process is known as Relevant Component Analysis

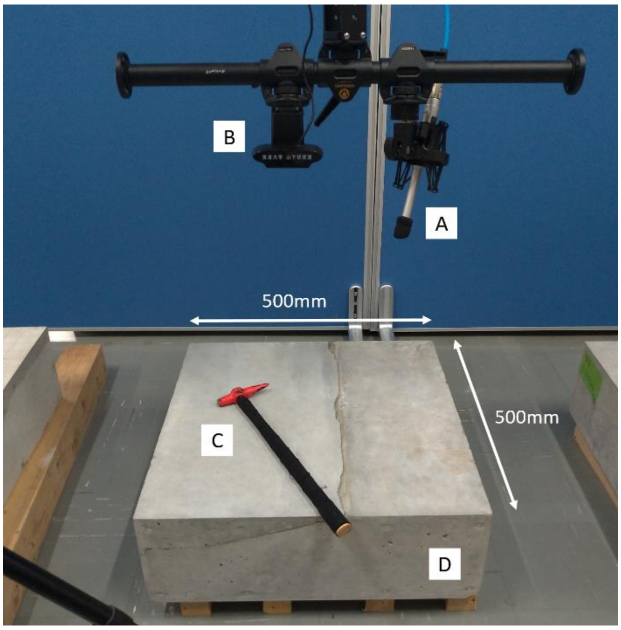

Fig. 3. Experimental setup with (A) a microphone, (B) a camera, (C) a hammer and (D) a concrete test block containing a man-made defect.

(RCA) [8]. Finally, K-Means is employed on the obtained feature space to separate hammering samples into defect and non-defect clusters.

3. Experimental Setup

The considered experimental setup is shown in Fig. 3. A concrete test block, containing a man-made defect, was used. This is the same concrete test block as the one considered in [7]. The crack propagated at an incident angle of 30 deg. from the surface for a length of 200 mm on the surface. 462 locations, 272 in non-defect area and 190 in defect area, where hit. The used hammer was a KTC UDHT-2, commonly used in inspection sites. Audio was recorded at 44.1 kHz with a Behringer ECM8000 and a Roland UA-25EX. MFCC were computed with 10 coefficients. Hammering location was detected using a Logitech HD Pro C920 using color tracking of the red hammer head. While there are several methods available for crack detection and tracing, crack detection was conducted manually so clustering performance evaluation could be conducted independently of crack detection performance.

$d_{\min}$ was set to about 20 cm, a usual minimum defect size. The Rand index, a common measure of performance for clustering in the range [0,1], was used to measure performance. The closer the Rand index is to 1, the better is the method.

Table 1: Average query success rate over 100 runs with 30 queries.

Method	Average query success rate
[7]	0.50
Simple filtering	0.62
Proposed method	**0.67**

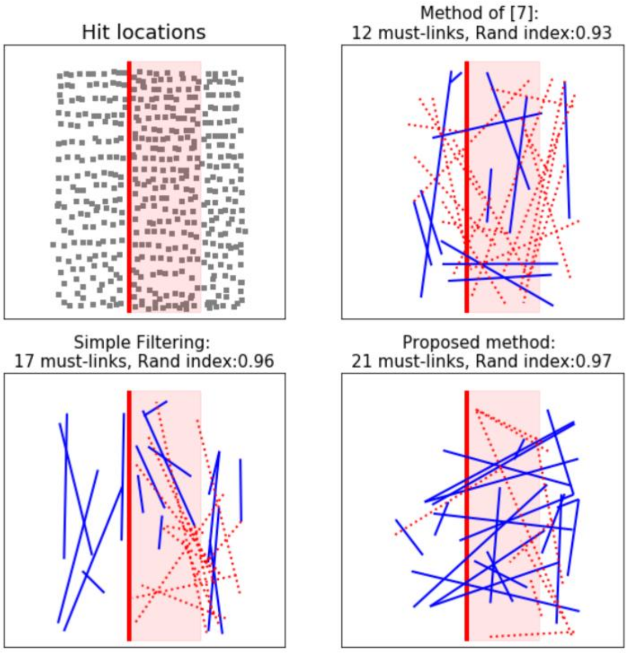

Fig. 4. Query examples: light red area show defect area, thick red line illustrates crack position, blue lines and dashed red lines represent queried must-links and cannot-links, respectively.

4. Results and Discussions

In Table 1 are reported the average query success rate, defined as the ratio of obtained must-links over total amount of queries, over 100 runs. Three methods were considered for comparison: the method of [7] using random query, a simple filtering forbidding query of pairs across cracks and the proposed method with query exclusion area.

It can be seen that without filtering, i.e., random selection of sample pairs to query the user on, the amount of effective must-links available for the RCA transformation is quite low: at an average query success rate of 0.50, only half of the queries resulted in the obtention of a must-link contributing to RCA. This also means that half of the queries to the user where meaningless. A simple filtering does allow for increased average query success rate at 0.62. Finally, the proposed

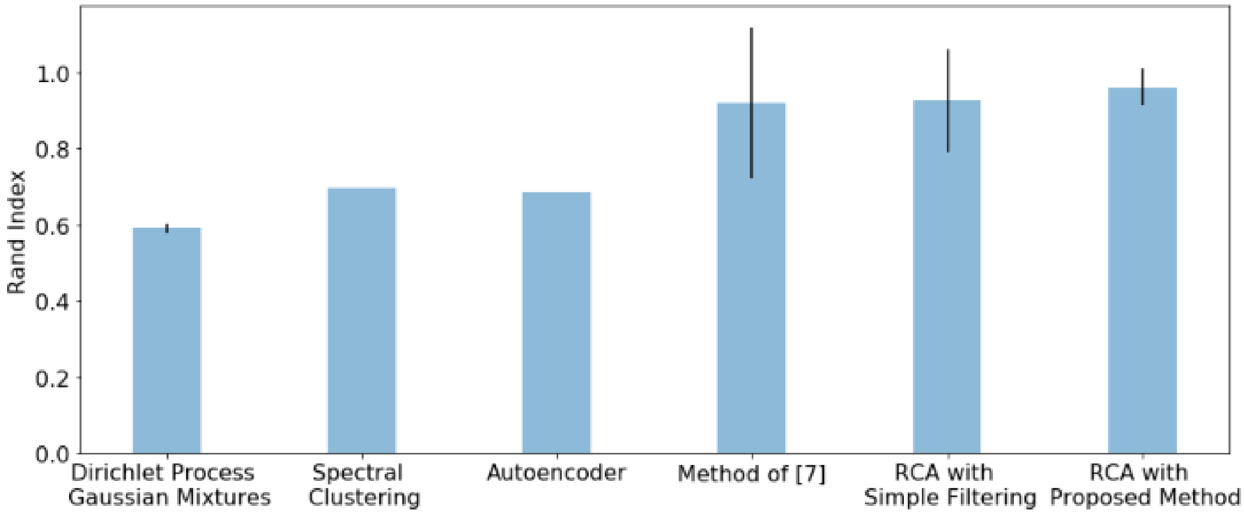

Fig. 5. Average performance over 100 runs with 30 queries. Error bars correspond to two standard deviations.

method yields much higher amounts of effective must-links from queries, at 0.67 average query success.

In Fig. 4 is shown an example of queries obtained under the considered methods. It can be noted that random queries effectively results in about half of queries being cannot-links, 12 out of 30 in this example, resulting in a Rand index of 0.92. Simple filtering yielded 17 must-links out of 30 queries, obtaining a Rand index of 0.96. The proposed method managed to obtain 21 must-links and a Rand index of 0.97. Aside the additional must-links obtained by our proposed method, one issue on this particular test block is the presence of non-defect areas on both sides of the crack: simple filtering prevents obtaining must-links between those two areas, unlike our proposed method. Such must-links are playing an important role in the metric learning process since they indicate that those two non-defect areas should be similar to each another, even if they are distant in both the physical and MFCC feature space.

In Fig. 5 are reported the average performance over 100 runs obtained by several methods. State-of-the-art machine learning methods such as Dirichlet Process Gaussian Mixtures [9], Spectral Clustering using the Euclidian distance and Autoencoder (K-Means over the feature space outputted by a Single Layer Perception with Gaussian noise added in training) were considered in addition to the method of [7], RCA with simple filtering and RCA with the proposed method.

Spectral Clustering reduces the feature space's dimension to the number of clusters, 2 in this case. This reduction might have been too drastic and negatively

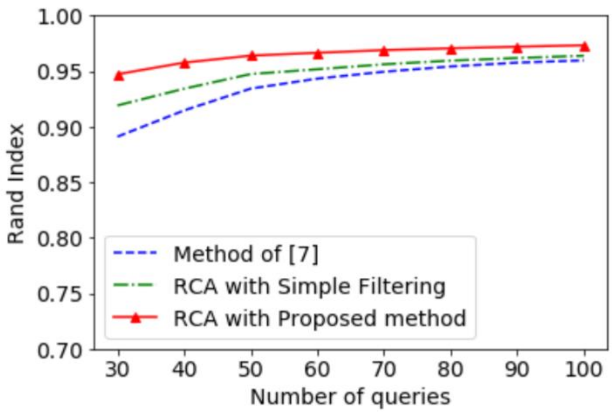

Fig. 6. Evolution of the clustering performance for the method of [7], RCA with Simple Filtering and RCA with the proposed method for varying number of allowed queries. Average of 20 runs per number of queries reported.

affected performance. The reported Autoencoder in Fig. 5 was a Single Layer Perceptron. Deeper network architectures returned lower performance. This is certainly due to such networks being more suited for non-linear encodings, while hammering class data has been reported to be linearly separable [4]. The relatively small dataset size certainly compounded the issue. Dirichlet Process Gaussian Mixtures selected a number of clusters with the stick breaking method and it was noticed that large numbers of clusters were created in most cases. This certainly explains its disappointing performance on this particular dataset.

The performance of [7], shown in Fig. 4, is much lower than reported previously since the query success ratio was not assumed to be perfect. It still showed good performance and outputted better clusterings than the methods mentioned earlier.

A simple filtering achieved a slight increase in average performance and slight reduction in standard deviation. Finally, the proposed method, yielding much higher amounts of effective must-links from queries, allowed RCA to provide much better average results as well as increased consistency.

Finally, in Fig. 6 is reported the average performance of weakly supervised methods for increasing amounts of queries. The lowest number of queries considered was 30, since RCA does requires at least the same number of must-links as the dimensionality of the data in order to obtain an inversible matrix in (3). For the three considered methods, the more queries were allowed, the better the performance was. It is worth noting that the advantage enabled by filtering queries is still apparent, even for high number of queries. RCA being basically a whitening process, once the major axes of the transformation are computed, additional must-links would serve as fine-tuning. This explains why the performance increase is not linear with the number of queries and also why filtering processes have lesser impact with high number of queries, since the additional must-link gain is less important. However, queries being a burden on the human user, it can be easily thought that low number of queries would be the practical case in the vast majority of applications and from this perspective, our proposed method is beneficial.

5. Conclusion

In this paper was proposed a method to filter queries for weakly supervised clustering using visual information for defect detection in concrete structures to increase efficiency and performance. By forbidding the consideration of pairs of samples located across the near vicinity of visible cracks, the effective amounts of must-links obtained notably increased for the same amount of allowed queries, increasing the efficiency of human involvement and, ultimately, performance.

In the future, we would like to expand further this work to encompass active queries to the human user, i.e., select the most relevant pairs for queries so as to obtain better outputs from RCA.

Acknowledgements

This work was supported in part by JSPS KAKENHI Grant Number 19J12391 and by the Japan Construction Information Center (JACIC) foundation. The authors also thank Prof. Hiromitsu Fujii. Chiba Institute of Technology and the Institute of Technology, Tokyu Construction Co., Ltd. for their cooperation.

参考文献

[1] BBC News, “Japan Sasago tunnel collapse killed nine. [web page],” https://www.bbc.com/news/world-asia-20576492, 2012.

[2] Henderson, Dion, and Costley, “Acoustic inspection of concrete bridge decks,” in Nondestructive Evaluation Techniques for Aging Infrastructures & Manufacturing, International Society for Optics and Photonics, 1999, pp. 219-227.

[3] Zhang, Harichandran, and Ramuhalli, “An automatic impact-based delamination detection system for concrete bridge decks,” in NDT&E International, Elsevier, 2012, vol. 45, pp. 120-127.

[4] Fujii, Yamashita, and Asama, “Defect detection with estimation of material condition using ensemble learning for hammering test,” in Proceedings of the International Conference on Robotics and Automation. IEEE, 2016, pp. 3847-3854.

[5] Louhi Kasahara, Fujii, Yamashita, and Asama, “Unsupervised Learning Approach to Automation of Hammering Test Using Topological Information,” in ROBOMECH Journal, 2017, vol. 4, no. 13, pp. 1-10.

[6] Louhi Kasahara, Fujii, Yamashita, and Asama, “Fuzzy clustering of spatially relevant acoustic data for defect detection,” in Robotics and Automation Letters, vol. 3, no. 3. IEEE, 2018, pp. 2616-2623.

[7] Louhi Kasahara, Fujii, Yamashita and Asama: "Weakly supervised approach to defect detection in concrete structures using hammering test,” Proceedings of the IEEE Global Conference on Consumer Electronics. IEEE, 2019, pp. 997-998.

[8] Bar-Hillel, Hertz, Shental, and Weinshall, “Learning a mahalanobis metric from equivalence constraints,” in Journal of Machine Learning Research, vol. 6, 2005, pp. 937-965.

[9] Blei and Jordan, “Variational inference for dirichlet process mixtures,” in Bayesian Analysis, vol. 1, 2006, pp. 121-143.