Fuzzy Clustering of Spatially Relevant Acoustic Data for Defect Detection

Jun Younes Louhi Kasahara, Hiromitsu Fujii, Atsushi Yamashita, and Hajime Asama

***Abstract*—Efficient diagnosis of concrete structures is a growing issue in modern societies where concrete is an omnipresent material. The hammering test is a traditional nondestructive testing method that has been employed in the field for a long time and for which automation is highly desirable. The problem consists in determining from the sound returned after a hammer strike on a structure's surface if there is a defect beneath or not. In this letter, we present an unsupervised learning approach for the automation of hammering test. First, mean shift is used with Mel-frequency cepstrum coefficients at various parameter values in order to find a stable mode configuration. Then, the corresponding peaks are used to obtain seeds for spatial fuzzy c-means to cluster hammering samples while combining audio and position information. Experiments have been conducted in indoor artificial environment on concrete blocks containing man-made defects. Results showed the effectiveness and robustness of the proposed solution on detecting different types and number of defects. Our approach showed promising performance also in tests performed in outdoor environment using a fully automated hammering system.**

***Index Terms*—Robotics in construction, robot audition.**

I. Introduction

CONCRETE is a material omni-present in modern societies and heavily used in construction, notably for social infrastructures such as bridges and tunnels. As it is the case with any material, concrete can be subject to damage by time and environmental factors. Therefore, diagnosis for defects in those structures is a paramount issue to guarantee user safety. There are several inspection methods for concrete such as ultrasonic CT and Computer Vision, however, for large scale in depth inspection, the hammering test in one of the most used methods. A human operator hits the surface of the structure with an inspection hammer, as shown in Fig. 1, and uses the returned sound to infer if the hit spot is a defect or not. This traditional method is popular for its non-destructive nature as well as for not requiring heavy and/or precise tools for execution, making it suitable for systematic inspection of big structures such as highways, bridges or tunnels. However, despite it apparent ease of use, the hammering test requires a skilled operator to correctly analyze hammering sounds. Other than making the diagnosis subjective, the growing number of structure to test demands a more efficient alternative [1]. Therefore, automation of the hammering test is highly desirable.

Fig. 1. Hammering test conducted by a professional: a hammer is used to hit the surface of the concrete structure and the returned sound used to assess whether a defect lies beneath the surface or not.

Besides direct methods such as the work in [2], where several testing methods were fused in a 3D visualization program, previous works in this field mainly used machine learning and can be distinguished between supervised and unsupervised learning approaches.

In [3] authors based their work on another yet similar acoustic non-destructive testing method for concrete structures which consists in dragging chains across the structure's surface. Physical noise reduction was conducted by sound-proofing around the chains and Linear Prediction Coefficients were used for detection. A modified Independent Component Analysis was used to separate noise from hammering sound and to train a Radial Basis Function neural network based on Mel-Frequency Cepstrum Coefficients (MFCCs) to classify hammering samples in [4]. In [5] the authors used ensemble learning techniques with time-frequency analysis in order to achieve a classifier able to detect defects and classify them according to their depth from the surface. This work was later expanded to encompass detection of crack propagation direction beneath the surface in [6]. These approaches have yielded remarkable results. However, the main drawback of these supervised learning approaches reside in their reliance on training sets. Those training sets, having a large influence on performance, are difficult to obtain,

Manuscript received October 11, 2017; accepted March 5, 2018. Date of publication March 28, 2018; date of current version April 11, 2018. This letter was recommended for publication by Associate Editor I. Rekleitis and Editor W. K. Chung upon evaluation of the reviewers' comments. This work was supported in part by the Council for Science, Technology and Innovation, "Cross-ministerial Strategic Innovation Promotion Program (SIP), Infrastructure Maintenance, Renovation, and Management" (funding agency: NEDO), in part by Japan Society for the Promotion of Science KAKENHI under Grant JP16H06680, and in part by the Institute of Technology, Tokyu Construction Co., Ltd. *(Corresponding author: Jun Younes Louhi Kasahara.)*

The authors are with the Department of Precision Engineering, Graduate School of Engineering, University of Tokyo, Tokyo 113-8656, Japan (e-mail: louhi@robot.t.u-tokyo.ac.jp; fujii@robot.t.u-tokyo.ac.jp; yamashita@robot.t.u-tokyo.ac.jp; asama@robot.t.u-tokyo.ac.jp).

Digital Object Identifier 10.1109/LRA.2018.2820178

especially in the case of concrete structures, due to logistical and legal reasons: data collection requires structures to be temporarily put out of service and there are legal issues if defects are found. Furthermore, due to many factors such as chemical composition, humidity and temperature during the hardening phase, concrete is a material with high variability. This results in hammering sounds being different from one structure to another, i.e., training sets being structure-specific.

An unsupervised learning approach based on rough clustering of hammering sounds' Fourier spectrum by k-means to obtain a model for the non-defect hammering sound was devised in [7]. That model was then used as a reference to conduct diagnosis using a correlation metric and thresholding. Despite yielding promising results in laboratory and field environment, the threshold selection by user at the final step hindered practicability and applicability for some cases. In our previous work, spatial fuzzy c-means and MFCCs were employed [8]. The proposed methodology showed high performance in the case of single and multiple delaminations. However, the fixed number of clusters restricted cases where this method was applicable. Indeed, in real-world conditions, the number and types of defects can vary greatly between structures. Another shortcoming was the random seeding process for fuzzy c-means: initial conditions, i.e., seeds, heavily influence the final clustering output. Therefore, a step to estimate the number of clusters on the inspected structure and an adequate seeding process are necessary.

In this letter, an offline unsupervised learning method for detection of defects in concrete structures, able to estimate the proper number of clusters present in the dataset and conduct clustering by discriminating defects, is presented. The letter is organized as follows: an overview of our proposed method is presented in Section II. Details on the estimation of number of clusters through mean shift are reported in Section III. Section IV describes the clustering of hammering samples based on audio and position data by spatial fuzzy c-means. Experimental tests and results performed in indoor artificial environment and outdoor on a mock tunnel are reported in Section V. Discussion is conducted in Section VI and conclusion are drawn in Section VII.

II. Overview of Proposed Method

Humans operators conduct hammering by usually hitting the structure's surface on multiple adjacent points and in quick succession: it can be thought that a hammering sample is not analyzed alone but as part of a group, i.e., operators would collect a certain amount of samples and look for samples that "do not sound like the others". Therefore, we choose to incorporate the notion of spatial autocorrelation in our approach. This can be roughly translated as "samples being located physically close together are more likely to belong to the same group" [9]. Since defects are localized and compact on a concrete structure, this concept fits well in this particular context. Humans' ability to conduct the hammering test also indicates that the human ear's perception range is good, or at least acceptable, to discriminate defect hammering samples. To emulate their performance, MFCCs are used as sound feature for hammering samples. This feature is devised to simulate human hearing and is popular in the field of speech recognition as well as other related fields such as music information retrieval [10], [11].

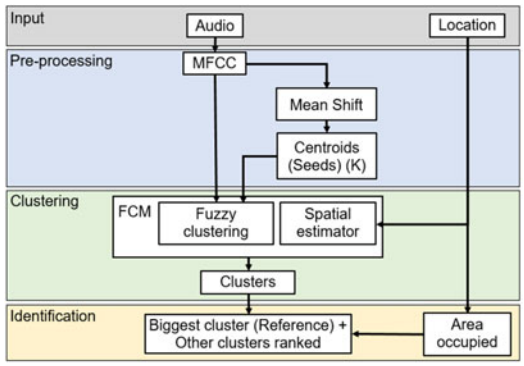

Fig. 2. Flowchart of proposed method.

Spatial fuzzy c-means was shown to have sufficient discriminative ability to differentiate hammering samples [8]. Furthermore, it is a staple in several fields and its simplicity allows for easy traceability of results.

However, the number of cluster K and seeds need to be chosen carefully. For this reason, mean shift is used to find stable mode configurations, i.e., cluster number, and the corresponding peaks in the probability function distribution are used as seeds for spatial fuzzy c-means. Mean shift only requires the tuning of one single parameter, namely bandwidth. Selection of this parameter is not a trivial task and heavily influences the output [12]: a low value of bandwidth tends to over-segment the data while higher values tend to merge clusters. To find an adequate value of bandwidth and successfully estimate K, stable mode configurations are searched by parsing the bandwidth parameter.

A flowchart of our proposed method is shown in Fig. 2. We assume that all hammering data has been collected and available prior to analysis.

Our contributions can be summarized as follows:

- We explore an unsupervised method for this issue.
- Previous work used a pre-determinated fixed number of clusters. Our proposed method has the capability of choosing a value of K adapted to the considered dataset.
- In addition to the determination of K, appropriate seed selection for clustering is conducted.

III. Seeding Using Mean Shift

Mean shift is a mode-seeking algorithm first proposed by [13]. It is appealing due to its non-parametric nature and has been popular in Computer Vision and Image Processing.

Mean shift considers that samples from a dataset are drawn from sources, most often modelled as Gaussian distributions. In the estimation of K, this assumption is accepted and an appropriate value of K and corresponding cluster centers are retrieved. Then, the proper cluster is conducted by spatial fuzzy c-means using mean shift's output. Spatial fuzzy c-means, contrary to mean shift, does not assume any model on data distribution. Furthermore, physical positions of samples are included in the analysis.

A. Feature Extraction

Considering a dataset of hammering samples $D = \mathscr{X}_1, \ldots, \mathscr{X}_N$ comprised of N samples, a hammering sample $\mathscr{X}_i = \{\mathbf{l}_i, \mathbf{x}_i\}$ is defined as:

1) Spatial Information: A physical location, point of contact between the hammer head and the structure's surface, noted $\mathbf{l}_i = (l_i^x, l_i^y)$. The physical distance between two samples is the Euclidian distance, noted $\|\mathbf{l}_i - \mathbf{l}_j\|$.

2) Audio Information using MFCCs: MFCCs of the sound resulting from the impact of the hammer head to the structure's surface, noted $\mathbf{x}_i$.

A metric based on the sample Pearson correlation coefficient, as we devised in [7], [8], is used for audio feature distance. Given the MFCCs $\mathbf{a} = (a_l)_{[l=1,\ldots,M]}$ and $\mathbf{b} = (b_l)_{[l=1,\ldots,M]}$, $d(\mathbf{a}, \mathbf{b})$ is computed as in (1), with M being the dimension of MFCCs and $\overline{a}$ and $\overline{b}$ being the average coefficient of $\mathbf{a}$ and $\mathbf{b}$, respectively.

$$d(\mathbf{a}, \mathbf{b}) = \frac{1}{2}\left(1 - \frac{\sum_{l=1}^{M}[(a_l - \overline{a})(b_l - \overline{b})]}{\sqrt{\sum_{l=1}^{M}(a_l - \overline{a})^2}\sqrt{\sum_{l=1}^{M}(b_l - \overline{b})^2}}\right) \tag{1}$$

B. Cluster Pruning and Bandwidth Selection

Due to the presence of outliers, clustering algorithms tends to produce irrelevant clusters, only containing a very small number of samples. Those irrelevant clusters correspond to either cases of over-segmentation by mean shift for low values of bandwidth or outliers, in spatial or feature domain, that end up being attributed their own individual cluster. Mean shift being used here to estimate K prior to actual clustering by spatial fuzzy c-means, the creation of those irrelevant clusters is highly undesirable. Therefore, to increase robustness against outliers and to seek "true" meaningful mode configurations of the dataset, we conducted cluster pruning, i.e., clusters that contain less than a certain percentage of the considered dataset were eliminated.

In order to obtain an appropriate value for the bandwidth parameter, stable mode configurations obtained by mean shift are investigated [14]. The bandwidth is tested for a range of possible values and the number of clusters or underlying peaks in the distribution, before and after the pruning step, of the dataset is observed. This results in plateaus in the number of cluster which indicates stable and meaningful mode configurations. Since slight over segmentation is preferable to under segmentation, the first plateau where the pruned and non-pruned number of cluster match is taken as the appropriate number of clusters. For the bandwidth value itself, the value of the middle of that plateau is chosen.

C. Seeds Extraction From Mean Shift

Fuzzy c-means presents the weakness of requiring the number of cluster to be known beforehand. Furthermore, the initialization step or seeding process can heavily influence the final clustering, i.e., a bad seeding process can lead the algorithm to be stuck in a local minima. To solve these two issues, we use the output of mean shift: to find in a data-driven approach the appropriate number of clusters for the considered dataset and to provide adequate seeds for fuzzy c-means in the form of centroids resulting from the clustering by mean shift.

IV. Clustering Using Spatial Fuzzy C-Means

Fuzzy c-means is a clustering algorithm widely used in many fields such as Computer Vision, more specifically for image segmentation [9].

A. Update Rule

1) Audio Feature Update Rule: The regular fuzzy c-means update is first conducted in MFCCs space using the correlation distance defined in (1). For each sample $\mathscr{X}_i$ toward each cluster center $\mathbf{c}_j$, fuzzy membership coefficient u_{ij} expresses how strongly the sample belongs to a cluster. With $(\mathbf{c}_j)_{[j=1,\ldots,K]}$ the cluster centroids and m a parameter controlling the fuzziness of the system, the update rule is conducted as in (2):

$$u_{ij} = \frac{1}{\sum_{l=1}^{N}\frac{d(\mathbf{x}_l, \mathbf{c}_i)}{d(\mathbf{x}_j, \mathbf{c}_j)}^{2/(m-1)}} \tag{2}$$

2) Spatial Feature Update Rule: Defects and non-defect areas are compact, e.g., a single non-defect sample physically surrounded by defect samples is likely to be a classification error, given that samples are physically close enough. In order to reduce these errors, first a spatial neighbourhood $NB(\mathscr{X}_i)$ for each sample $\mathscr{X}_i$ is defined as in (3): a disc of radius γ based on Euclidian distance over samples' physical locations. Then, based on this neighbourhood, a spatial estimator h_{ij} , the "expected" fuzzy membership of sample $\mathscr{X}_i$ for cluster j based on other samples in proximity, is defined as the average of corresponding fuzzy memberships, as shown in (4), with $|NB(\mathscr{X}_i)|$ being the number of neighbours for sample $\mathscr{X}_i$. Compared to simply conduct a smoothing step after a regular fuzzy c-means, spatial fuzzy c-means enables to find a balance between these inputs of different nature, namely position and audio information, by conducting smoothing at each iteration and combining it with clustering in audio space.

$$NB(\mathscr{X}_i) = (\mathscr{X}_i \in D \mid \|\mathbf{l}_i - \mathbf{l}_j\| \leq \gamma) \tag{3}$$

$$h_{ij} = \frac{1}{|NB(\mathscr{X}_i)|}\sum_{k \in NB(\mathscr{X}_i)} u_{kj} \tag{4}$$

Finally, with p and q weighting exponents on each fuzzy components, an additional spatial update is conducted as in (5):

$$u_{ij} \rightarrow \frac{u_{ij}^p h_{ij}^q}{\sum_k u_{kj}^p h_{kj}^q} \tag{5}$$

The centroid update rule remains unchanged from the regular fuzzy c-means. Conversion to a crisp clustering is done by maximum membership: each hammering sample is assigned to the cluster with the highest fuzzy coefficient.

B. Cluster Identification

Our proposed method does not distinguish between defect types, i.e., every type of defect are regrouped under the label *defect*.

Here, a couple of different cases arise. If mean shift seeding step returns a single cluster, it is assumed that no defects are present in the considered dataset. There is the possibility that the whole dataset is composed of defects, however, assuming that the defects do not occupy the majority of the structure and the dataset covers a large enough area, this possibility can be reasonably put aside: if the inspection target was damaged to such extend, precise diagnosis would not be required, a simple visual inspection would suffice. If two clusters are identified, the method used in [7], [8] is applied: a weight for each cluster based on hammering samples' physical distance to nearest neighbour r_i, as shown in (6), is computed, as in (7), to find the cluster occupying the largest area of the structure $\mathbf{c}_{\text{non-defect}}$, as in (8).

$$r_i = \frac{1}{2} * \min_{\forall \mathscr{X}_j \in D} \|\mathbf{l}_i - \mathbf{l}_j\| \tag{6}$$

$$w_{\mathbf{c}_k} = \sum_{\mathscr{X}_i \in \mathbf{c}_j} r_i^2 \tag{7}$$

$$\mathbf{c}_{\text{non-defect}} = \mathbf{c}_{\arg\max_k (w_{\mathbf{c}_k})} \tag{8}$$

In the last case of more than two clusters, as previously weights for each cluster are computed and the largest cluster is labelled as non-defective with its center $\mathbf{c}_{\text{ref}}$. Then, other clusters are ranked based on the correlation distance from that cluster and labelled non-defective if they fall under a threshold T, defective otherwise, as shown in (9), with $label_{\mathbf{c}_i}$ the label attributed to the j-th cluster.

$$label_{\mathbf{c}_i} = \begin{cases} \text{defective} & \text{if } d(\mathbf{c}_i, \mathbf{c}_{\text{ref}}) \geq T \\ \text{non-defective} & \text{otherwise} \end{cases} \tag{9}$$

V. Experiments

A. Devices and Settings

The used experimental setup is illustrated in Fig. 3. Location of a hammering sample was a 2D position, achieved by color-tracking the hammer head. The used hammer was a KTC UDHT-2 (length 380 mm, weight 160 g, head diameter 16 mm), commonly used in hammering test by professionals.

Sound was recorded at 44.1 kHz using a Behringer ECM8000 condenser microphone coupled with a Roland UA25EX sound board and a laptop PC for data analysis. Fourier spectrum were obtained by Fast Fourier Transform (FFT) using a window size of 1024 and then MFCCs were computed using 26 triangular filters after zero mean and unit variance normalization. Only the first 10 coefficients, excluding the zero coefficient, were then kept. The zero coefficient represents the average log-energy of the input signal and does not interest us here since zero mean normalizations are effected. Higher coefficients represents fast changes in the Mel filterbanks energies and are known to degrade recognition capabilities [15]. p and q were set to unity, $m = 2$ and γ was set so that every sample would have at least one neighbour. A laptop with a Core i7 at 2.60 Ghz was used in the experiments and only a few seconds was needed for analysis.

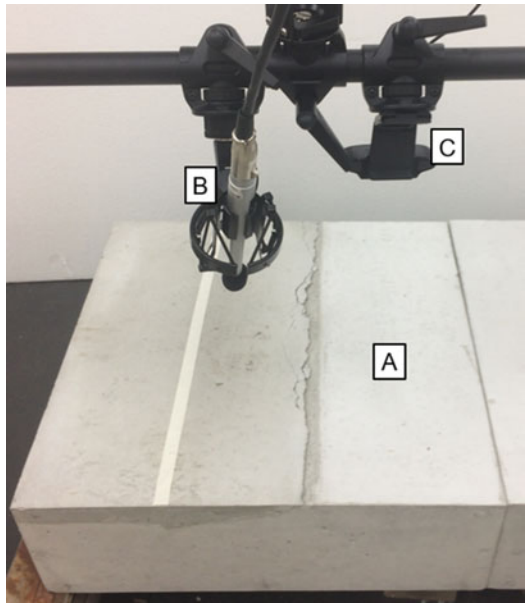

Fig. 3. Experimental setup in laboratory conditions showing a hammer on top of (a) a concrete test block, (b)microphone and (c) camera.

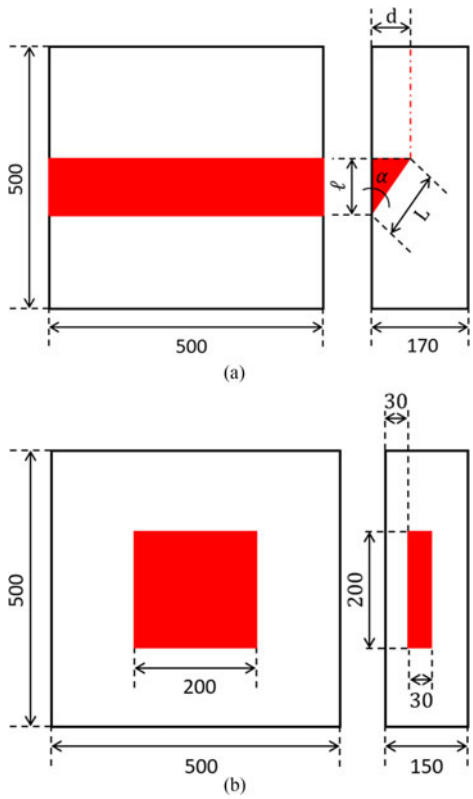

Fig. 4. Schematic of concrete test blocks: red area shows defects. Dimensions in mm. (a) Generic schematic of the delamination type concrete test block. (b) Schematic of the void type concrete test block.

B. Using Concrete Test Blocks

For testing purposes, concrete test blocks containing man-made defects were made, as shown in Figs. 4 and 5. The corresponding graphs used for bandwidth selection are shown in Fig. 6. The outputs returned for each case using normalized

Fig. 5. Concrete test blocks containing man-made defects used in our experiments. The red areas shows ground truth of defects. (a) Case 1: single delamination. (b) Case 2: dual delamination. (c) Case 3: delamination & void. (d) Case 4: no defects.

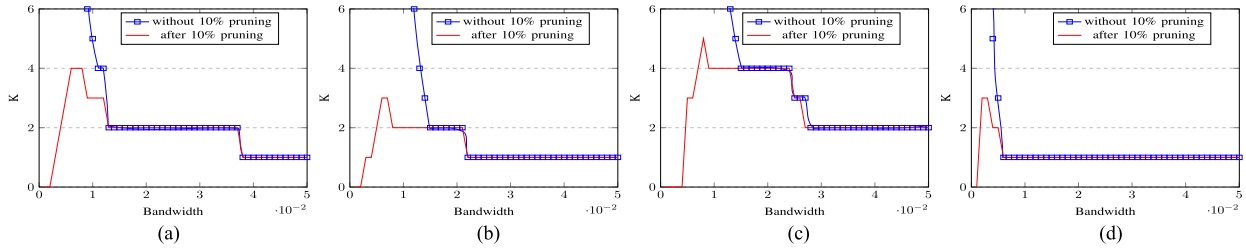

Fig. 6. Evolution of the number of clusters found by mean shift for various values of bandwidth parameter. (a) Case 1. (b) Case 2. (c) Case 3. (d) Case 4.

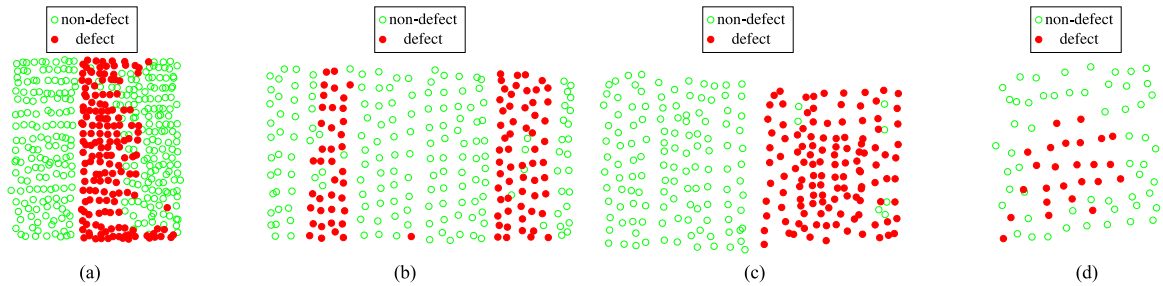

Fig. 7. Results of normalized spectral clustering on the considered cases. (a) Result of normalized spectral clustering on Case 1. (b) Result of normalized spectral clustering on Case 2. (c) Result of normalized spectral clustering on Case 3. (d) Result of normalized spectral clustering on Case 4.

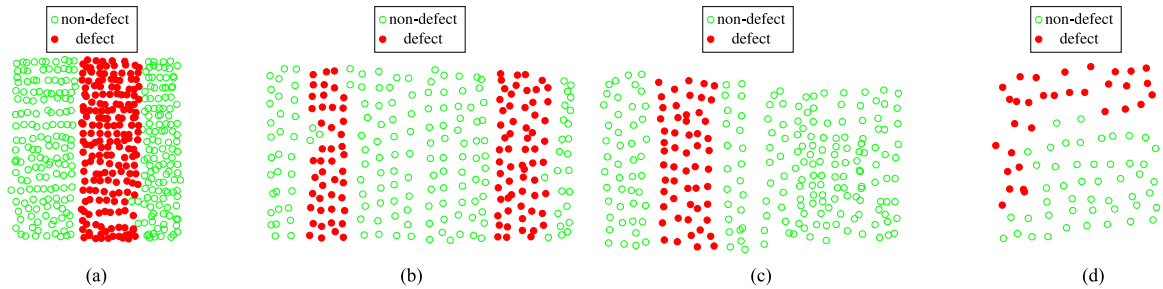

Fig. 8. Results of [8] on the considered cases. (a) Result of [8] on Case 1. (b) Result of [8] on Case 2. (c) Result of [8] on Case 3. (d) Result of [8] on Case 4.

spectral clustering [16] (with Fourier spectrum, Euclidian distance metric and setting $K = 2$), our previous work [8] and our proposed method were examined and are shown in Figs. 7–9, respectively: red dots and green circles represent hammering samples classified as defects and non-defects respectively, location of samples correspond to Fig. 5. Performance in each case was quantified under the values of precision, recall, and accuracy and are presented in Table I. Hammering action was here conducted by human while trying to follow a regular grid pattern as much as possible with the hammer strike motion: this allows most samples to have the same number of neighbours and thus to establish a fair spatial estimator.

1) Case 1. Single Delamination: The tested concrete block is shown in Fig. 5(a) and contained a single delamination. The dataset is composed of 462 samples: 272 non-defects and 190 defects. Referring back to Fig. 4(a), $\alpha = 30$ deg., $l = 200$ mm, $d = 115.5$ mm and $L = 230.9$ mm. In Fig. 6(a) is shown the evolution of cluster number K with the bandwidth. It can be noticed that K stabilizes first at a plateau of $K = 2$. The value of 0.025 was chosen for this case. As shown in Fig. 7(a), normalized spectral clustering returned an overall good result (Table I). Some misclassified samples on the left side, where the defect runs deeper, can be noted. The work of [8] and our proposed method returned similar outputs.

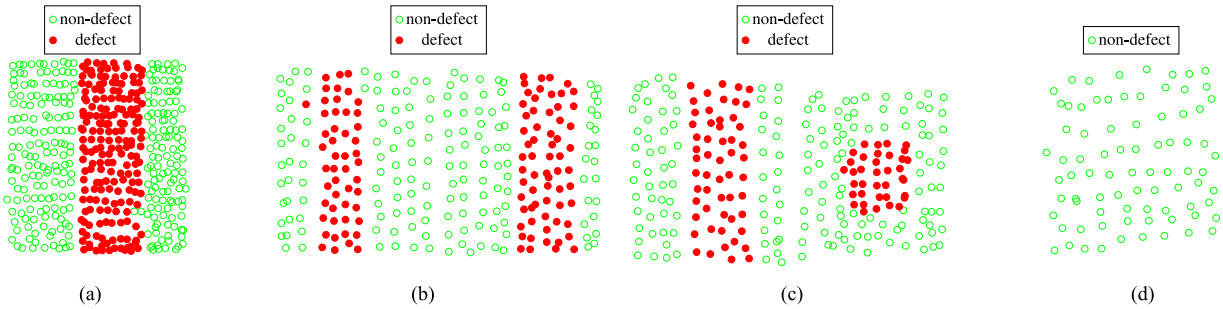

Fig. 9. Results of our proposed method on the considered cases. (a) Result of proposed method on Case 1. (b) Result of proposed method on Case 2. (c) Result of proposed method on Case 3. (d) Result of proposed method on Case 4.

TABLE I
PERFORMANCE OF VARIOUS METHODS ON EXPERIMENTS CONDUCTED USING CONCRETE TEST BLOCKS

	Case 1: single delamination			Case 2: dual delamination			Case 3: delamination and void			Case 4: no defects		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy
Spectral clustering	93.2%	78.9%	89.0%	99.0%	85.2%	93.3%	32.8%	42.1%	46.1%	-	-	68.8%
[8]	100.0%	99.0%	99.6%	100.0%	95.2%	95.2%	100.0%	57.9%	84.3%	-	-	60.0%
Proposed method	**100.0%**	**98.9%**	**99.6%**	**99.1%**	**98.9%**	**98.9%**	**98.9%**	**95.8%**	**98.0%**	-	-	**100.0%**

Values of precision, recall and accuracy, usually used to measure a classifier's performance, were computed.

2) Case 2. Dual Delamination: The corresponding concrete test blocks are shown in Fig. 5(b). To simulate the presence of multiple delaminations in the tested area, two concrete blocks containing each a single delamination were put together. This dataset is composed of 270 samples: 155 non-defects and 115 defects. For the left side block $\alpha = 15$ deg., $d = 40$ mm, $l = 149.3$ mm, and $L = 154.5$ mm. For the right side block $\alpha = 15$ deg., $l = 200$ mm, $d = 53.6$ mm and $L = 207.1$ mm. The number of clusters in the bandwidth selection process stabilized at a value of $K = 2$ and this plateau is meaningful i.e., matches the curve of cluster number after cluster pruning at 10% of the dataset (Fig. 6(b)). Results shows that our proposed method and the approach in [8] returned better performance than using normalized spectral clustering, as illustrated in Figs. 7(b)–(b). Compared to the work in [8], our proposed method yielded a slightly lower precision for a little gain in recall and accuracy, as indicated in Table I.

3) Case 3. Delamination and Void: The corresponding concrete test blocks are shown in Fig. 5(c). Here, one concrete block containing a delamination (left side block, the same as the one used on the right in case 2) and another one containing a polystyrene cuboid to simulate a void (right side block, schematic shown in Fig. 4(b)) were put together. This aimed to simulate the co-existence of defects of different nature in a tested area. This dataset is composed of 254 samples: 159 non-defects and 95 defects. According to Fig. 6(c), the setting $K = 4$ seemed appropriate. The result outputted by normalized spectral clustering mostly managed to differentiate the two blocks, without being able to discern the defects embedded in each of them. This resulted in poor performance, as indicated in Fig. 7(c) and Table I. Using MFCCs and spatial fuzzy c-means, it is interesting to note that the method in [8] managed to perfectly discern the delamination but failed to detect the void defect due to the setting $K = 2$, as shown in Fig. 8(c). Our proposed method on the other hand achieved to discern both defects, Fig. 9(c), with high performance values of precision, recall and accuracy (Table I).

4) Case 4. No Defects: The corresponding concrete test block is shown in Fig. 5(d). This particular concrete block did not contain any defects and the dataset is composed of 80 non-defect samples. Fig. 6(d) shows that the number of clusters dives straight to a stable and meaningful at 10% plateau at $K = 1$, strongly suggesting that the data is composed of a single cluster. Both normalized spectral clustering and the method in [8] divided the data into two clusters, as shown in Figs. 7(d) and (d). The difference in outputs can be explained by the different audio feature used (Fourier spectrum and MFCCs).

To summarize, our proposed method was successfully able to replicate the results presented in [8] in cases 1 and 2, with comparable performance. Moreover, it showed its ability to correctly estimate the number of clusters and conduct clustering in case 3 and 4, where the previous approaches failed to return satisfying results.

C. In Field Conditions

The performance of our proposed method was tested in field conditions using a mock tunnel, shown in Fig. 10(a). It is located outside and in scale with tunnels currently in service. Also, in opposition with the concrete test blocks presented earlier, defects in this tunnel occurred naturally. Therefore, it can be assumed that this setting is identical to actual inspection sites for hammering with the benefits of availability for experimentation. The defects here being natural, we were limited in the types of defects available for testing. Ground truth was obtained here by meticulous inspection by a seasoned professional. Since there was no time constraints, the results of this inspection can be considered superior of what is usually conducted regularly in the field and thus very close to the actual ground truth. Due to inaccessibility of true ground truth and limited number of areas

(a) (b)

Fig. 10. Field conditions experimental setting. (a) Mock tunnel: located outdoors and in real scale, this tunnel exhibits naturally occurred defects and closely matches real-life field conditions. (b) Hammering module mounted on Variable Guide Frame: this system allows automatic hammering and data collection anywhere inside a tunnel [17]–[19].

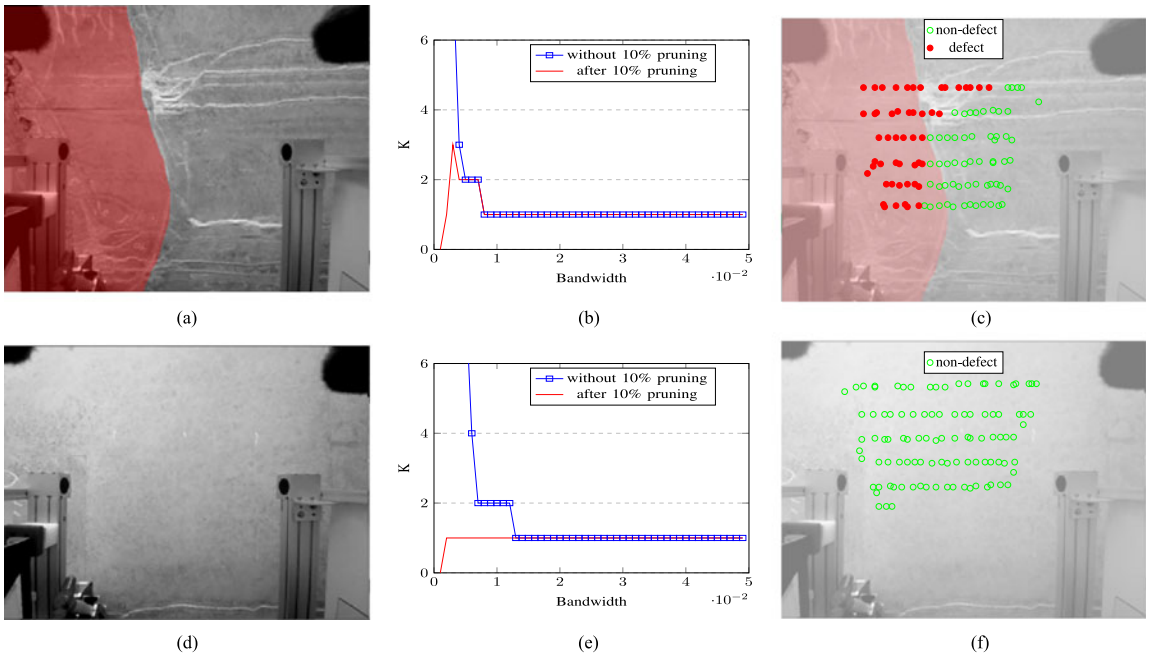

Fig. 11. Results of our proposed method on portions of the mock tunnel. (a) Area 1: delamination. The red area shows ground truth of defects. (b) Evolution of number of cluster with bandwidth. (c) Result of proposed method on Area 1. (d) Area 2: no defects. (e) Evolution of number of cluster with bandwidth. (f) Result of proposed method on Area 2.

of interest in the tunnel, we restricted ourselves to qualitative analysis of the few obtained results.

Hammering was effected and data collected using an automated hammering module mounted on a system called *Variable Guide Frame*, as illustrated in Fig. 10(b) [17]–[19]. This setup allows to conduct hammering virtually anywhere on the surface of the tunnel, in a fashion similar to human operators by careful design of the hammer swing motion to reproduce human-like hammering sounds.

1) Area 1. Delamination: The first area considered on the mock tunnel contains a delamination. Furthermore, this area has been subject to rainwater leakage, resulting in the white lines visible in Fig. 11(a). Fig. 11(b) shows that a plateau exists for $K = 2$. The curve after cluster pruning confirms that this mode configuration is meaningful. Our proposed method's output in this case, shown in Fig. 11(c), shows that most of the samples seem to have been correctly classified.

2) Area 2. No Defects: Here, a stable mode configuration of $K = 2$ for bandwidth around the value 0.001 would seem to exist. However, it appears that this is due to outliers as that particular plateau does not appear after cluster pruning. Therefore, the correct value of $K = 1$ can be selected.

In summary, our proposed method was able to return satisfying results in field conditions by successfully estimating the cluster number in the cases of delamination and absence of defects using a fully automated hammering system.

VI. Discussion

Case 1 and 2 in artificial indoor environment showed that our proposed method matched the performance of our previous work [8]. As MFCCs contains most of the meaningful information in the first coefficients, dimension reduction does not seem to have impacted significantly performance. Furthermore,

in these cases our proposed method also indicated that the correct number of cluster was 2, as the user speculated in [8]. In case 3 consisting of two blocks with delamination and void type defects, the approach in [8] failed. This is certainly due to the fixed value of $K = 2$: the dataset could be more easily split between samples of each block than between defects and non-defects. Blocks in this case were from different batches: this would explain the difference in number of clusters compared to case 2, i.e., the non-defect areas of each blocks sounded different enough to be each attributed a cluster. The result obtained using [8] shows that the delamination defect was successfully discriminated but the void defect was grouped in the non-defect cluster. Our proposed method on the other hand successfully estimated a correct number of cluster along with seeds that yielded enough "clustering resolution" for defects in each block to be given a cluster each: 2 clusters for non-defect samples (one per block), 1 cluster for delamination and 1 cluster for void. This case shows best the limitation of a fixed number of clusters used in previous works and the need of an selection step for K. In case 4, where there was no defects, i.e., a single cluster, direct use of the method described in [8] could not provide satisfying results: fuzzy c-means class would forcibly divide the dataset to try to match the value of K given by the user. Analysis of meaningful modes by mean shift here enables to overcome one of fuzzy c-means' fundamental weakness. However, the problem of the current identification step is obvious here: in the presence of a single cluster, there is no real guarantee that this cluster is non-defective.

Analysis of meaningful mode configurations for the estimation of K was shown to be valid in field conditions as well, although the plateau were less obvious. This is certainly due to the presence of noise that is bound to be recorded along hammering sound in an outdoor setting. Some samples on the upper portion in Area 1 were mislabeled as defects: this could be attributed to the poor surface condition caused by rainwater leakage that resulted in degraded hammering sounds.

VII. Conclusion

In this letter, we proposed an unsupervised learning approach for accurate and robust detection of defects in concrete structures. The proposed method allows for proper selection of number of clusters in considered hammering datasets and seeding for clustering using spatial fuzzy c-means. Experiments conducted in both indoor and outdoor environments demonstrated the validity of the proposed solution. Results showed that our method exhibit excellent performances in case of single and multiple delamination detection. Moreover, it yielded remarkable results also in the presence of multiple defects of different nature as well as in the case where no defects were present. Furthermore, results of the experiments performed in outdoor environment are also very promising.

In the future, we would like to continue development of this approach, especially in the identification phase after clustering, in order to avoid the need for user input and to add the capability not only to discriminate defects versus non-defects but also classification of types of defects, i.e., achieve multi-class classification as a final output instead of the current bi-class classification. This could be achieved by more careful analysis of the fuzzy clustering prior to the crisp clustering conversion through defuzzification. Bringing other sources of information, such as visual information, in order to conduct multi-modal inspection is also in consideration. Furthermore, with the recent advances in automatic hammering modules, easier hammering data acquisition can be expected in the near future: given enough data, incorporation of deep learning techniques could yield extremely promising results.

References

[1] R. Takada and N. Oishi, "Priority issues of infrastructure inspection and maintenance robot: A part of COCN 2012 project disaster response robot and its operational system," in *Proc. Humanitarian Tech. Conf.*, 2013, pp. 166–171.

[2] J. Kim, N. Gucunski, T. H. Duong, and K. Dinh, "Three-dimensional visualization and presentation of bridge deck condition based on multiple NDE data," *J. Infrastructure Syst.*, vol. 23, no. 3, 2016, Art. no. B4016012.

[3] M. E. Henderson, G. N. Dion, and R. D. Costley, "Acoustic inspection of concrete bridge decks," in *Nondestructive Evaluation Techniques for Aging Infrastructures & Manufacturing*. Bellingham, U.K.: SPIE, 1999, pp. 219–227.

[4] G. Zhang, R. S. Harichandran, and P. Ramuhalli, "An automatic impact-based delamination detection system for concrete bridge decks," *NDT E Int.*, vol. 45, no. 1, pp. 120–127, 2012.

[5] H. Fujii, A. Yamashita, and H. Asama, "Defect detection with estimation of material condition using ensemble learning for hammering test," in *Proc. Int. Conf. Robot. Autom.*, 2016, pp. 3847–3854.

[6] J. Im, H. Fujii, A. Yamashita, and H. Asama, "Multi-modal diagnostic method for detection of concrete crack direction using light-section method and hammering test," in *Proc. Ubiquitous Robots Ambient Intell.*, 2017, pp. 922–927.

[7] J.Y. Louhi Kasahara, H. Fujii, A. Yamashita, and H. Asama, "Unsupervised learning approach to automation of hammering test using topological information," *ROBOMECH J.*, vol. 4, no. 1, pp. 13–23, 2017.

[8] J.Y. Louhi Kasahara, H. Fujii, A. Yamashita, and H. Asama, "Clustering of spatially relevant audio data using Mel-frequency cepstrum for diagnosis of concrete structure by hammering test," in *Proc. Int. Symp. Syst. Integr.*, 2017.

[9] K. S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Comput. Med. Imaging Graph.*, vol. 30, no. 1, pp. 9–15, 2006.

[10] T. Ganchev, N. Fakotakis, and G. Kokkinakis, "Comparative evaluation of various MFCC implementations on the speaker verification task," in *Proc. Int. Conf. Speech Comput.*, 2005, vol. 1, pp. 191–194.

[11] M. Müller, *Information Retrieval for Music and Motion*. Berlin, Germany: Springer, 2007.

[12] D. Comaniciu, "An algorithm for data-driven bandwidth selection," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 25, no. 2, pp. 281–288, Feb. 2003.

[13] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 1, pp. 32–40, Jan. 1975.

[14] Comaniciu and Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, 2002.

[15] K. S. Rao and S. G. Koolagudi, *Robust Emotion Recognition Using Spectral and Prosodic Features*. New York, NY, USA: Springer Science & Business Media, 2013.

[16] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[17] Y. Takahashi, S. Nakamura, Y. Ogawa, and T. Satoh, "Velocity control mechanism of the under-actuated hammering robot for gravity compensation," in *Proc. 34th Int. Symp. Autom. Robot. Construction*, 2017, pp. 446–451.

[18] S. Nakamura, Y. Takahashi, D. Inoue, and T. Ueno, "The variable guide frame vehicle for tunnel inspection," in *Proc. 34th Int. Symp. Autom. Robot. Construction*, 2017, pp. 671–674.

[19] F. Inoue, K. Soonsu, T. Uchiyama, S. Nakamura, and Y. Yanagihara, "Shape control of variable guide frame for tunnel wall inspection," in *Proc. Int. Symp. Autom. Robot. Construction*, 2017, pp. 675–682.