

Article

# Acoustic Inspection of Concrete Structures Using Active Weak Supervision and Visual Information

Jun Younes Louhi Kasahara \*<sup>†</sup>, Atsushi Yamashita <sup>†</sup> and Hajime Asama <sup>†</sup>

Department of Precision Engineering, The University of Tokyo, Tokyo 113-8656, Japan;  
yamashita@robot.t.u-tokyo.ac.jp (A.Y.); asama@robot.t.u-tokyo.ac.jp (H.A.)

\* Correspondence: louhi@robot.t.u-tokyo.ac.jp

† These authors contributed equally to this work.

Received: 27 December 2019; Accepted: 21 January 2020; Published: 23 January 2020



**Abstract:** Concrete structures are featured heavily in most modern societies. In recent years, the need to inspect those structures has been a growing concern and the automation of inspection methods is highly demanded. Acoustic methods such as the hammering test are one of the most popular non-destructive testing methods for this task. In this paper, an approach to defect detection in concrete structures with active weak supervision and visual information is proposed. Based on audio and position information, pairs of samples are actively queried to a user on their similarity. Those are used to transform the feature space into a favorable one, in a weakly supervised fashion, for clustering defect and non-defect samples, reinforced by position information. Experiments conducted in both laboratory conditions and in field conditions proved the effectiveness of the proposed method.

**Keywords:** acoustic inspection; defect detection; clustering; weak supervision; active query

---

## 1. Introduction

Concrete structures are featured heavily in most modern societies. This is especially true for social infrastructures such as tunnels, bridges, and highways. Such large scale structures require regular inspection to prevent the propagation of damages caused by various factors such as rain and wind. The importance of such inspection work was underlined in recent catastrophic failures such as the collapse of the Sasago tunnel in Japan [1] or, more recently, the collapse of the Morandi bridge in Italy [2].

Inspection of large scale concrete structures is paramount to ensure the safety of their users and several methods exist for this task such as using capacitive transducers [3], microwave imaging [4], or embedded piezoceramic transducers [5].

One popular method for the inspection of such structures is the hammering test. It consists of hitting the surface of the tested area with a hammer and using the returned impact sound to assess the presence of defects under the surface, as illustrated in Figure 1. This method is extremely popular due to its simplicity and non-destructive nature.

There are two factors complicating the inspection of concrete structures. Currently, inspection work is predominantly conducted by veteran human inspectors and there is a shortage of such skilled workers. On the other hand, the population of concrete structures reaching ages where inspection is critically required is ever increasing. Therefore, the automation of inspection methods such as the hammering test is highly demanded.

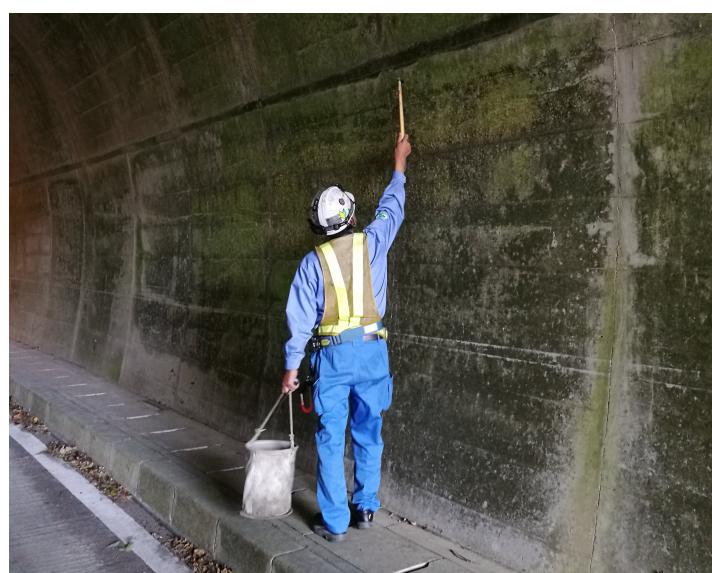
While there are some previous works employing direct methods [6], most of them used machine learning approaches. In [7], the sound of dragging chains across concrete surfaces was used with Linear Prediction Coefficients. In [8], Independent Component Analysis was used along a Radial Basis Function Neural Network. In [9], Ensemble Learning was used with time-frequency analysis. In [10],

Partial Least Squares Regression was used to predict the gap in concrete-metal composite structures. While achieving high performance, such supervised learning approaches have an inherent dependency on the available training data. For concrete structures, such training data are neither available nor easy to produce.

The other end of the machine learning spectrum is called unsupervised learning. In [11], Mode Decomposition, Principal Component Analysis, and Independent Component Analysis were used for concrete bridge inspection. In [12,13], clustering methods were used. Those bypassed the issue of training data but involved strong priors in their design to compensate for the absence of training data, for which no mitigation would be possible.

Weakly supervised methods, also known as semi-supervised methods, are a mix of supervised and unsupervised methods. Such methods aim to achieve the best of both worlds. From the point of view of concrete structure inspection, such approaches are also extremely attractive since human involvement can be maintained during the actual inspection process. The work in [14] proposed an initial framework for weakly supervised inspection of concrete structures, later expanded by the addition of visual information in [15]. However, weak supervision was assumed to be provided by the human user on randomly chosen pairs of samples. However, this makes for weak supervision of inconsistent quality. Weak supervision quality greatly affects the performance of the resulting defect detection and, in most cases, it is possible to actively query the user for specific weak supervision.

Therefore, in the present paper, an active weakly supervised method reinforced by visual information for the inspection of concrete structures using acoustic methods is proposed.



**Figure 1.** Hammering test conducted by a professional on the upper wall portion of a tunnel: only a simple hammer is needed, thus the popularity of this non-destructive testing method. However, there is the need of skilled operators to correctly differentiate hammering sounds and given the great population of structures in need of testing, automation is actively demanded. In addition, since it relies heavily on the operator's skills, the final result remains subjective.

## 2. Our Novelty and Overview of Proposed Method

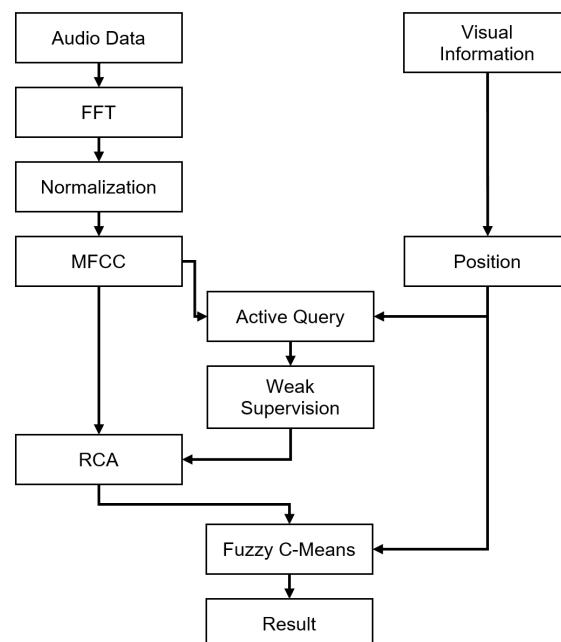
The method of Louhi Kasahara et al. [13] was unsupervised, i.e., did not require supervision of any form by a human user, and used position information in order to reinforce the clustering in audio feature space. The proposed method is a weakly supervised method and therefore requires a human user to provide weak supervision on each considered dataset of hammering samples. The proposed method uses position information to both reinforce clustering in the feature space obtained by weak supervision and in the active query process, i.e., select pairs of samples to query the human user on.

The methods of Louhi Kasahara et al. [14] and Louhi Kasahara et al. [15] were weakly supervised. Louhi Kasahara et al. [14] only used weak supervision, as opposed to Louhi Kasahara et al. [15] and the proposed method that use position information to complement weak supervision. The method of Louhi Kasahara et al. [15] assumed that the human user would randomly select pairs of samples for weak supervision. This resulted in weak supervision of inconsistent quality and, thus, inconsistent defect detection performance. The proposed method actively queries the user: hammering samples are first processed and a selection process to decide which pairs of samples the human user should provide weak supervision on is proposed.

Our contributions can be summarized as follows:

- We explore an active weakly supervised method for the issue of defect detection in concrete structures.
- Previous work passively waited for a user to provide weak supervision and had no control over its quality. Our proposed method actively queries the user to increase the probability of obtaining quality weak supervision.
- Visual information is also employed to use the position of hammering samples in the active query as well as in the analysis in order to reinforce weak supervision results.

An overview of the proposed method is shown in Figure 2. During the hammering process, audio data are recorded and visual information is used to record the location of each hammering sample, i.e., the position on the tested structure where the hammer head contacted the concrete surface. Audio data are processed into Fourier spectrum using Fast Fourier Transform (FFT), normalized and finally converted into Mel-Frequency Cepstrum Coefficients (MFCC). Active query, based on both audio and visual information, is used to obtain quality weak supervision on MFCC samples and Relevant Component Analysis (RCA) [16] is used to learn an appropriate metric for discriminating defect hammering samples. Finally, Fuzzy C-Means, a fuzzy clustering algorithm, is used to discriminate defect and non-defect hammering samples.



**Figure 2.** Overview of proposed method for active weakly supervised defect detection using hammering and visual information.

### 3. Preprocessing

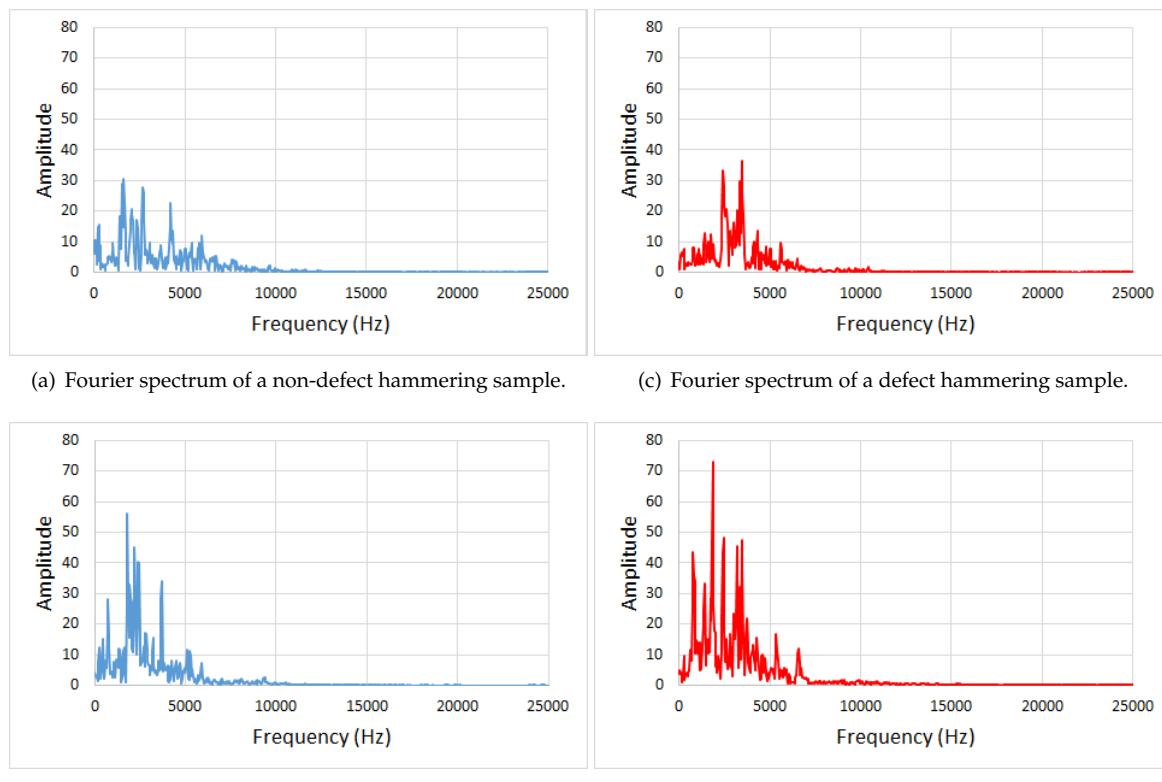
#### 3.1. Fourier Spectrum and Position Information

Audio-samples are initially time-series data and not really suited for analysis. It has been long established in audio signal processing that frequency analysis, i.e., analysis using the Fourier spectrum, is a more suitable approach. The Fourier spectrum of an audio signal can be obtained using the Fourier transform.

The Fourier transform consists of decomposing a time-series function, i.e., a signal, into the different frequencies it contains. The Fourier transform of a signal in the time domain is a function in the complex domain, with its absolute value representing the amount of each frequency present in the original signal and its complex argument representing the phase offset. The absolute value of the Fourier transform, designated as simply *Fourier spectrum* in the remaining of this paper, is used as initial feature vector for a sound sample. Given a sound sample defined by  $\mathbf{x} = (x_1, \dots, x_d)$ , its Fourier spectrum  $\mathbf{a} = (a_1, \dots, a_d)$  is defined as in Equation (1) using FFT, an algorithm computing the discrete Fourier transform.

$$a_j = \left| \sum_{l=1}^d x_l e^{-\frac{2\pi i}{d} j l} \right| \quad j = 1, \dots, d \quad (1)$$

For illustration, Fourier spectrums of a couple of non-defect and defect hammering samples are shown in Figure 3.



**Figure 3.** Fourier spectrums of hammering samples. It can be noted that features allowing discrimination of defect hammering samples are not obvious.

#### 3.2. Normalization

In the most general audio-based inspection setting, there is no assumption about the regularity of the input. In the case of the hammering test for concrete structures, there is no information about how much energy is used to hit the structure and to generate sounds. Especially in the case where a human

conducts the striking motion, as opposed to a robot, there are no guarantees that the same amount of force is used for every audio sample and therefore correct comparison among those audio samples cannot be conducted.

To conduct comparison between audio samples, with  $\bar{a}$  being the mean of the components of  $a_i$  as defined in Equation (2) and  $d$  the dimensionality of  $a$ , a normalization to zero mean and unit variance is conducted as in Equation (3) to obtain the normalized Fourier spectrum  $\tilde{a} = (\tilde{a}_1, \dots, \tilde{a}_d)$ :

$$\bar{a} = \frac{1}{d} \sum_{i=1}^d a_i, \quad (2)$$

$$\tilde{a}_i = \frac{a_i - \bar{a}}{\sqrt{\frac{\sum_{i=1}^d (a_i - \bar{a})^2}{d-2}}}. \quad (3)$$

### 3.3. Mel-Frequency Cepstrum Coefficients

The work presented in [12] showed the effectiveness of MFCC as feature vectors for hammering samples to emulate the human audition's ability to conduct the hammering test. MFCC is a hand-crafted feature vector originally built for speech recognition, devised with a strong understanding of how human beings perceive sounds. It is widely popular in the field of speech recognition as well as other related fields such as music information retrieval [17,18].

The main steps for building MFCC are as follows:

1. Calculate the periodogram estimate of the power spectrum.
2. Apply Mel filterbanks, i.e., ensemble of filters, to the power spectrum and sum the energy in each filter.
3. Calculate the logarithm of all filterbank energies.
4. Calculate the Discrete Cosine Transform (DCT) of the log filterbank energies.

The first step is to compute the periodogram estimate of the power spectrum as in Equation (4):

$$P_i = \frac{1}{N_{window}} |\tilde{a}_i|^2. \quad (4)$$

Mel filterbanks are a set of  $N_{filter}$  triangular filters, usually between 20 to 40, equally spaced in the Mel scale presented in Equation (5), where  $f$  is the frequency in Hertz. The Mel scale is an empirical scale tuned to the previously mentioned sensitivity of the human cochlea.

$$M(f) = 1125 * \ln\left(1 + \frac{f}{700}\right). \quad (5)$$

This filterbank is applied to the periodogram estimate of the power spectrum. This provides  $N_{filter}$  energy values. Finally, the DCT of the logarithm of those energy values are calculated. The result is what are called MFCC.

For the sake of clarity, in the remainder of this paper, the MFCC of a hammering sample is denoted using simply  $x$ . Therefore, the dataset  $D = \{\mathcal{X}_i\}_{i \in [1 \dots N]}$  is composed of hammering samples  $\mathcal{X}_i = \{x_i, l_i\}$ , with  $l_i$  corresponding to the hit location on the tested structure.

## 4. Active Weakly Supervised Metric Learning

### 4.1. Active Query

The performance of weakly supervised methods is conditioned by the quality of the provided weak supervision. In the proposed method, RCA, a weakly supervised metric learning method, is employed. RCA learns an adapted metric for the clustering task at hand based on the weak supervision provided by a human user, i.e., hints on how the feature space should be. Concretely, weak supervision

consists of what are called *constraints* or, more precisely, *must-links*, pairs of samples indicated as similar through querying the user. If those hints are distributed badly in the feature space, the transformation to the adapted feature space cannot be achieved with satisfying performance. In our previous work, it was assumed that the human user would randomly choose pairs of samples to provide as constraints. In the proposed method, the pairs of samples to query the user are actively selected in an effort to provide better performance.

The default metric, usually the Euclidian distance, is not satisfactory for discriminating defect hammering samples in the MFCC feature space. However, the shortcomings of a non-suited metric mostly appear over medium to short ranges, as local structures of the data are usually easier to discern than global ones. More concretely, if two samples are very similar to each other, their distance across various metrics would remain small and the difference between metrics would be relatively small. On the other hand, the correct metric should be able to capture accurately the distance between dissimilar samples. In terms of weak supervision, this means must-links between samples separated by a big default metric value are more likely to contribute to a meaningful change in the computed new metric.

Moreover, one particular aspect of hammering samples is that they are composed of audio and position data. This spatial correlation of hammering samples was shown to provide great help in defect detection performance [12]. Samples that are physically located near one another can be expected to usually have strong similarity since they resulted from hits of the hammer on similar concrete surfaces. The only case where this does not hold is for defect boundaries, where closely located samples do belong to different classes and are therefore dissimilar. This is not relevant for the scope of the proposed method since only similar pairs of samples are used as weak supervision. Therefore, it is desirable to obtain must-links for sample pairs that are located physically far from each other, i.e., weak supervision on distant pairs of samples for which the shortcomings of the default metric are most likely to be apparent.

To take into account the two concepts presented in the two previous paragraphs, we propose an active query scheme based on both distances in the audio space and physical space. Pairs of samples  $\{\mathcal{X}_i, \mathcal{X}_j\}$  are selected to be queried to a user based on the query selection probability  $P(i, j)$  defined as in Equation (6), where the first and second term of the numerator correspond to the default metric in MFCC and physical space, respectively, and the denominator is a normalizing term.

$$P(i, j) = \frac{|\mathbf{x}_i - \mathbf{x}_j| + |\mathbf{l}_i - \mathbf{l}_j|}{\sum_{(k,l), k < l} |\mathbf{x}_k - \mathbf{x}_l| + |\mathbf{l}_k - \mathbf{l}_l|}. \quad (6)$$

The proposed active query scheme is described in Algorithm 1.

---

**Algorithm 1:** Pseudo-code of proposed active query scheme.

---

**Data:** Dataset  $D$  of  $N$  samples  $\mathcal{X}_i$ , target number of must-links  $N_q$   
**Result:** Set of must-links  $\mathcal{M}$

```

for each sample pair  $\{\mathcal{X}_i, \mathcal{X}_j\}$  do
     $P(i, j) = \frac{|\mathbf{x}_i - \mathbf{x}_j| + |\mathbf{l}_i - \mathbf{l}_j|}{\sum_{(k,l), k < l} |\mathbf{x}_k - \mathbf{x}_l| + |\mathbf{l}_k - \mathbf{l}_l|}$ 
end
while  $|\mathcal{M}| < N_q$  do
    Randomly select a sample pair  $\{\mathcal{X}_i, \mathcal{X}_j\}$  based on  $P(i, j)$ 
    Query the user on the selected pair  $\{\mathcal{X}_i, \mathcal{X}_j\}$ 
    if The answer is positive then
        | Add the selected sample pair  $\{\mathcal{X}_i, \mathcal{X}_j\}$  to the set of must-links  $\mathcal{M}$ 
        Set selection probability of selected pair to null:
         $P(i, j) \rightarrow 0$ 
end

```

---

#### 4.2. Relevant Component Analysis

RCA is a metric learning clustering method proposed initially in [16]. Several variants exists [19–21] and recently several successful applications have been reported [22,23].

Based on weak supervision, RCA computes a linear transformation of the feature, where the clustering task would be easier. Using must-links, a process akin to whitening is conducted. Given  $N_{\text{chunklet}}$  chunklets  $\{\mathcal{M}_l\}_{l \in [1 \dots N_{\text{chunklet}}]}$ , which are must-links regrouped using their transitive property, with  $\hat{\mathbf{m}}_l$  being the mean of elements in  $\mathcal{M}_l$ , RCA can be divided into three steps:

1. For each chunklet, subtract its mean from each sample it contains.
2. Compute the covariance matrix  $\hat{\mathbf{C}}$  as in Equation (7), with  $N_{\text{total}}$  being the total number of elements contained in the chunklets.

$$\hat{\mathbf{C}} = \frac{1}{N_{\text{total}}} \sum_{j=1}^{N_{\text{chunklet}}} \sum_{\mathbf{x}_i \in \mathcal{M}_l} (\mathbf{x}_i - \hat{\mathbf{m}}_l)(\mathbf{x}_i - \hat{\mathbf{m}}_l)^T. \quad (7)$$

3. Compute the whitening transformation associated with this covariance matrix, i.e., the inverse square root of the covariance matrix, and apply it to the dataset as in Equation (8).

$$\mathbf{x}_{\text{new}} = \hat{\mathbf{C}}^{-1/2} \mathbf{x}. \quad (8)$$

### 5. Clustering with Position Information

Fuzzy C-Means is a suited clustering framework to incorporate spatial information along with the main data type [24] and has been successfully used with hammering data in [13]. It is a fuzzy clustering algorithm, meaning that samples belong to several clusters at the same time, with varying degrees expressed through *fuzzy membership coefficients*. Fuzzy C-Means in our proposed method is composed of two fuzzy membership coefficient update steps that are iterated until convergence has been reached. The algorithm is described in Algorithm 2.

#### 5.1. Audio Feature Update Rule

The first update is the regular Fuzzy C-Means update, conducted on MFCC features only. For each sample  $\mathcal{X}_i$  toward each cluster center  $\mathbf{c}_j$ , the corresponding fuzzy membership coefficient is noted  $u_{ij}$ . With  $\{\mathbf{c}_j\}_{j=1 \dots K}$ , the cluster centroids, and  $m$ , a parameter controlling the fuzziness of the system, the update rule is conducted as in Equation (9):

$$u_{ij} = \frac{1}{\sum_{l=1}^N \frac{d(\mathbf{x}_l, \mathbf{c}_i)^2 / (m-1)}{d(\mathbf{x}_j, \mathbf{c}_j)}}. \quad (9)$$

#### 5.2. Spatial Feature Update Rule

The compacity of defect and non-defect areas of the tested structure is expressed through the introduction of a spatial estimator, based on a spatial neighborhood  $NB(\mathcal{X}_i)$  for each sample  $\mathcal{X}_i$ . This is defined using the position information of hammering samples, as in Equation (10). A spatial estimator  $h_{ij}$  is then used to estimate the fuzzy membership coefficients of the considered sample based on its neighborhood, as in Equation (11), with  $|NB(\mathcal{X}_i)|$ , the number of neighbors for sample  $\mathcal{X}_i$ . This process enables keeping the same *hammering resolution* and offers a much more stable output compared to a simple smoothing step after clustering.

$$NB(\mathcal{X}_i) = (\mathcal{X}_i \in D \mid |\mathbf{l}_i - \mathbf{l}_j| \leq \gamma), \quad (10)$$

$$h_{ij} = \frac{1}{|NB(\mathcal{X}_i)|} \sum_{k \in NB(\mathcal{X}_i)} u_{kj}. \quad (11)$$

Combination of the spatial estimator and regular Fuzzy C-Means update is conducted as in Equation (12), with  $p$  and  $q$  weighting exponents on each fuzzy components:

$$u_{ij} \rightarrow \frac{u_{ij}^p h_{ij}^q}{\sum_k u_{kj}^p h_{kj}^q}. \quad (12)$$

The centroid update rule remains unchanged from the regular Fuzzy C-Means. Conversion to a crisp clustering is done by maximum membership. The whole process is summarized in Algorithm 2.

---

**Algorithm 2:** Pseudo algorithm of Fuzzy C-Means with Position Information.

---

**Data:** Dataset  $D$  of  $N$  samples  $\mathcal{X}_i$ , number of clusters  $K$   
**Result:** Partitioning of  $D$  into  $K$  fuzzy clusters expressed through fuzzy coefficients  $u_{ij}$

```

foreach hammering sample  $\mathcal{X}_i \in D$  do
    | Assign random fuzzy coefficients  $u_{ij}$ 
end
while termination criterion is not met do
    foreach cluster centroid  $\mathbf{c}_j$  do
        | Conduct centroid update
        |  $\mathbf{c}_j = \frac{\sum_i u_{ij}^m \mathbf{x}_i}{\sum_i u_{ij}^m}$ 
    end
    foreach hammering sample  $\mathcal{X}_i \in D$  do
        | Update in MFCC space
        |  $u_{ij} = \frac{1}{\sum_{l=1}^N d(\mathbf{x}_l, \mathbf{c}_i)^{2/(m-1)}}$ 
        | Build spatial estimator
        |  $h_{ij} = \frac{1}{|NB(\mathcal{X}_i)|} \sum_{k \in NB(\mathcal{X}_i)} u_{kj}$ 
        | Combine
        |  $u_{ij} \rightarrow \frac{u_{ij}^p h_{ij}^q}{\sum_k u_{kj}^p h_{kj}^q}$ 
    end
end

```

---

## 6. Experiments

### 6.1. Experimental Setup In Laboratory Conditions

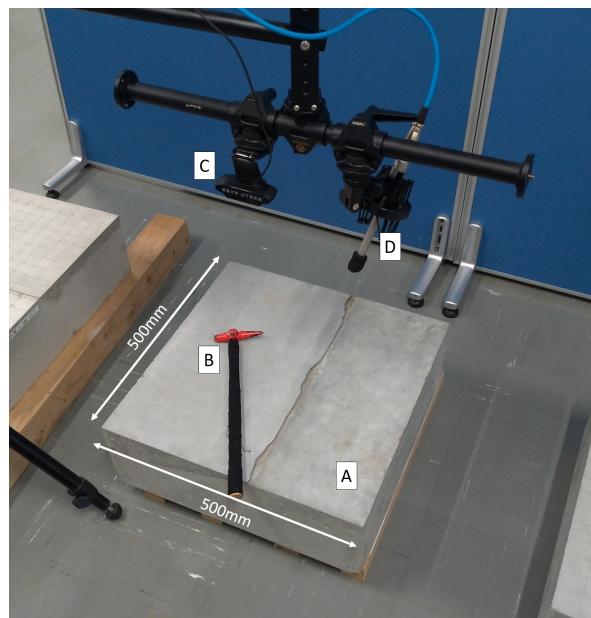
The used setup is illustrated in Figure 4 and experiments were conducted on concrete test blocks containing various man-made defects to simulate natural ones. Standard concrete mixture ratios used in Japanese tunnels were also used for those concrete test blocks. Depending on the mixture ratios of concrete, the sound of the hammering test can vary greatly. For each type of structure, strict standards define the mixture ratio of concrete to be used and human inspectors are trained on each specific type of structure. Our proposed method does not rely on training data and, therefore, assuming that the human user providing the weak supervision has been trained on the corresponding concrete mixture ratio, it is applicable to all concrete mixture ratios.

Test blocks were hit at the upper surface on several locations, once per location. The used hammer was a KTC UDHT-2 (head diameter 16 mm, length 380 mm, weight 160 g), commonly used in hammering test by professionals and sound was recorded at 44.1 kHz using a Behringer ECM8000 microphone coupled with a Roland UA-25EX soundboard and a laptop for data analysis. Fourier Spectrums of length 1024 were obtained using FFT and MFCC were computed with 10 coefficients. The hammer head was painted in red and the location of each hammering sample was obtained by

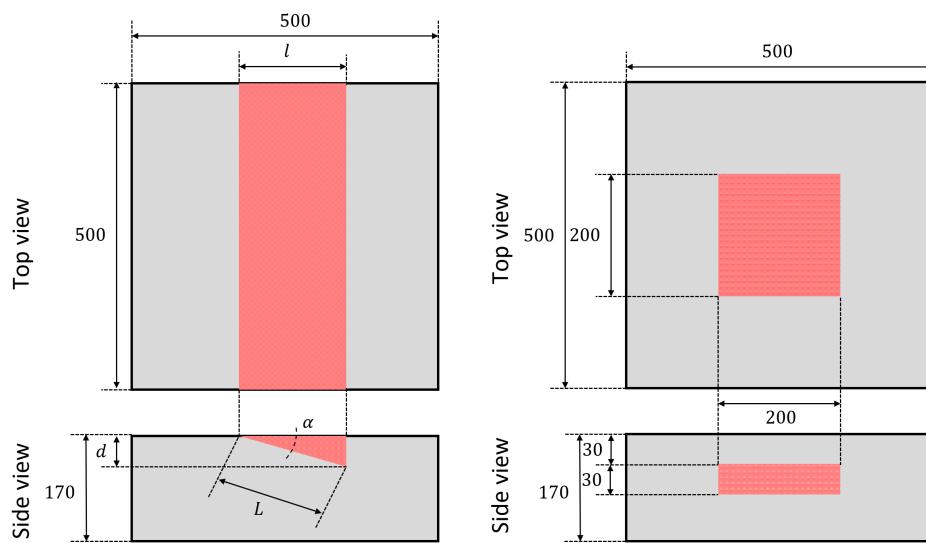
color-tracking the hammer head. In all experiments,  $p$  and  $q$  were set to the default value of 1, meaning equal contribution of audio data and visual information.

Figure 5 shows a generic schematic of the blocks used in experiments. Red areas correspond to defect areas. Three scenarios were considered:

- Case 1 corresponds to a situation where the tested area of a concrete structure contains a single delamination. The dataset is composed of 462 samples: 272 non-defects and 190 defects. Referring back to Figure 5,  $\alpha = 30$  deg.,  $l = 200$  mm,  $d = 115.5$  mm and  $L = 230.9$  mm. The setting  $K = 2$  was used.
- Case 2 aims 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. Referring back to Figure 5, 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 setting  $K = 2$  was used.
- In Case 3, one concrete block containing a delamination and another one containing a polystyrene cuboid to simulate a void were put together. This aims to simulate the co-existence of defects of different natures in a single tested area. This dataset is composed of 254 samples: 159 non-defects and 95 defects. The setting  $K = 4$  was used here since the two blocks were from different concrete batches.



**Figure 4.** Experimental setup: concrete test block (A); hammer with red head for position recognition by image processing (B); web camera (C); and microphone (D).



(a) Generic schematic of the delamination type concrete test block.  
(b) Schematic of the void type concrete test block.

**Figure 5.** Schematic of the concrete test blocks containing man-made defects. Compared to natural defects found in the field, such concrete test blocks have the advantage of having the ground truth, i.e., location of defects, known in advance, without the need to destroy them, due to precise and elaborate fabrication processes.

## 6.2. Experimental Setup In Field Conditions

Field experiments in actual concrete structures currently in service present big logistic and legal issues. Therefore, experiments were conducted in a mock tunnel, as shown in Figure 6a. Its structure, fabrication process, and scale match the profile of small one way tunnels commonly found in Japan. In outdoor conditions, the mock tunnel differs from the previously presented concrete test blocks in the nature of its defects: they occurred naturally. Therefore, this mock tunnel can be considered to very accurately match actual inspection sites, while presenting the advantage of availability. Since the defects were not man-made here, their number and type were limited.

Two areas of interest were found in the mock tunnel, Area 1 and Area 2, both corresponding to a ceiling portion where a delamination was found. Furthermore, Area 1 presented extensive surface damage caused by rainwater leakage. Here, the ground truth of defect was obtained by the help of a veteran human professional, with expert knowledge of the hammering test. Great care was given during his inspection of the mock tunnel and, therefore, this human inspection result can be considered to be superior to what is usually conducted in actual inspection sites, where time is often a constraint. The setting  $K = 2$  was used for both Area 1 and Area 2.

Furthermore, the hammering was conducted in this mock tunnel using an automated hammering module, mounted on a system named *Variable Guide Frame*. An illustration is provided in Figure 6b [25–27]. This particular setup can collect hammering samples virtually anywhere inside a tunnel and the hammering motion was designed to mimic those of human inspectors. As reported in [25], the hammering module was designed to output consistent hammering at 0.2 J using a hammer head weighting 133.8 g, i.e., a hammering force of about 1.5 N.



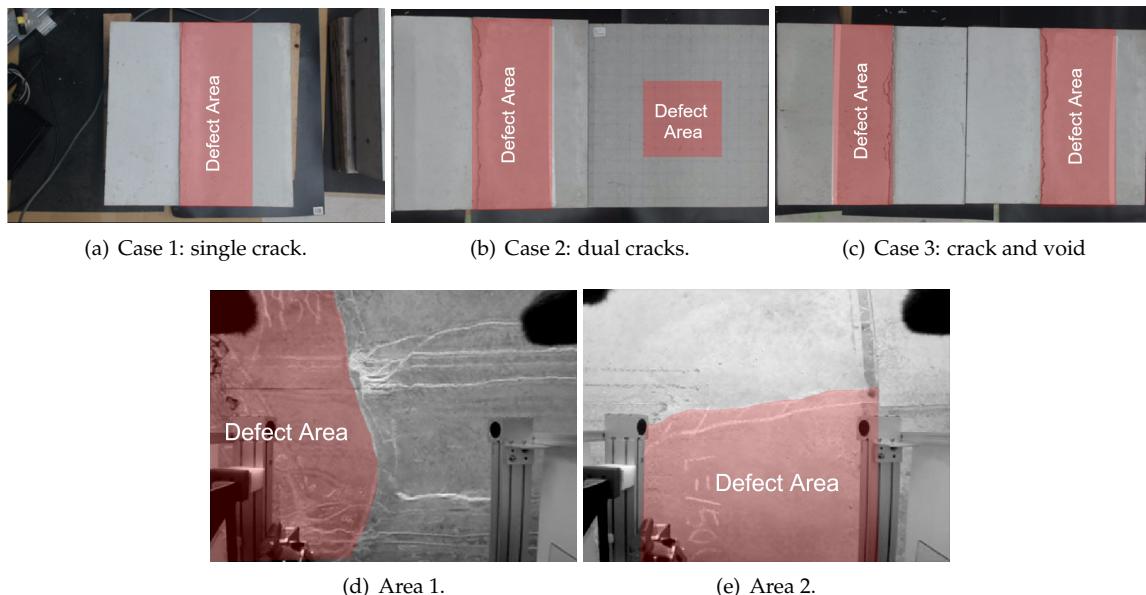
(a) Mock tunnel used in field experiments. Conditions as close as possible of actual inspection sites were realized due to its scale and naturally-occurred defects.

(b) Hammering module mounted on Variable Guide Frame. This setup allows automatic collection of hammering samples in any location inside of a tunnel.

**Figure 6.** Field experimental setup using a mock tunnel and hammering module [13].

## 7. Results and Discussions

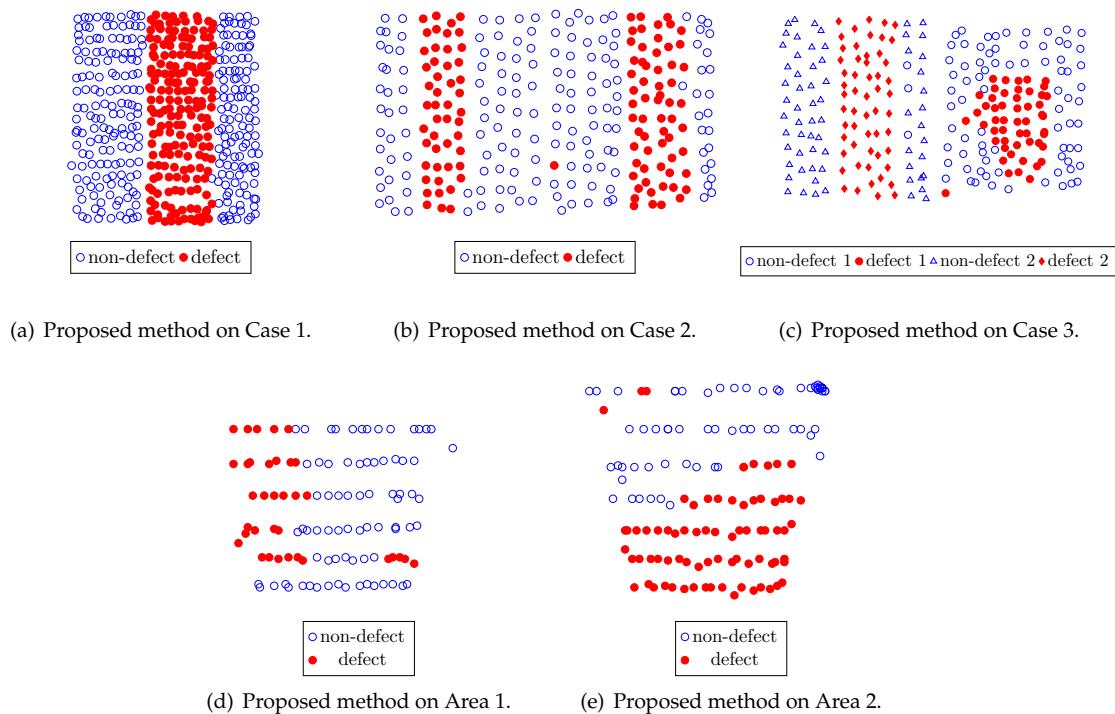
Figure 7 shows the tested concrete blocks and areas in the mock tunnel, with each defect area shown with a red overlay. Figure 8 illustrates the defect detection outputs using the proposed method. Figure 9 reports the performance obtained using the method of Louhi Kasahara et al. [14], RCA with the proposed active query method, the method of Louhi Kasahara et al. [15], and the proposed method.



**Figure 7.** Picture of the considered cases in laboratory conditions and in field conditions. Red areas indicate defect areas.

In Figure 8a,b, it can be seen that the separation defect/non-defect hammering samples by our proposed method is excellent, only missing a couple of samples. In Case 3, shown in Figure 8c, the output of the proposed method shows an overspill of the void defect area. However, the two distinct defects, as well as the two distinct blocks, were rather well discerned. Regarding Area 1 and Area 2, the proposed method outputted the results reported in Figure 8d,e. The overall result is good. Misclassified samples are mainly located on the edge of the target area, hinting at possible variations induced by the hammering module. The defect/non-defect borders were however well discerned and

the degraded surface condition of Area 1 did not seem to negatively affect the performance of our proposed method.



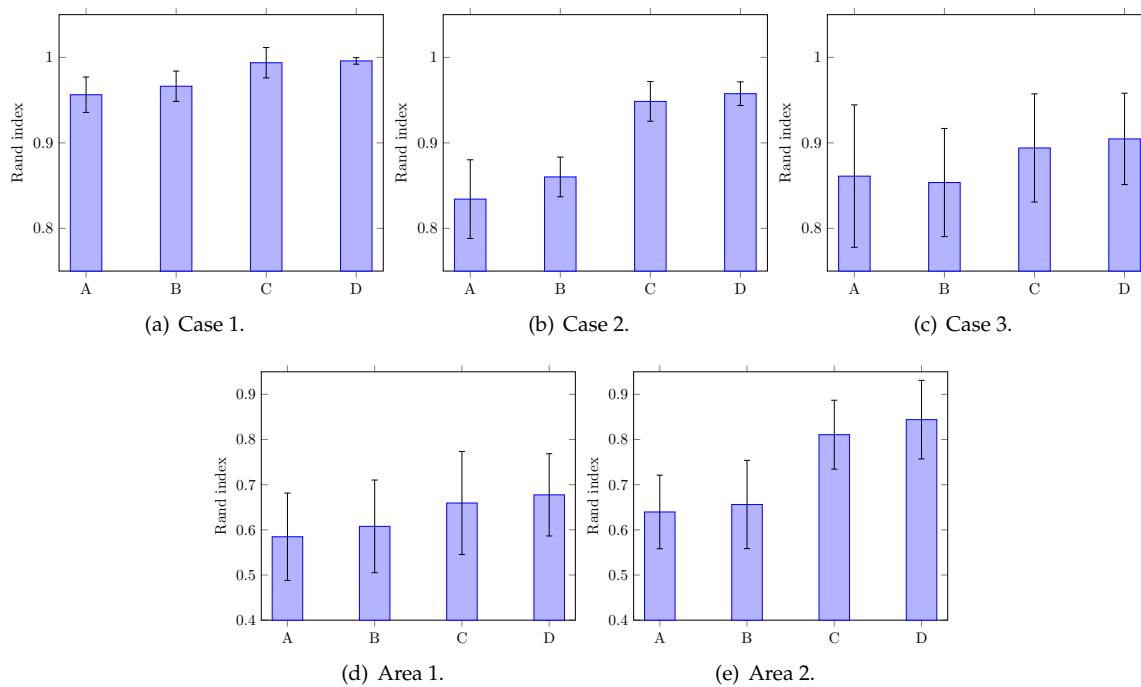
**Figure 8.** Example outputs of proposed method.

RCA with the proposed active query scheme should be put in direct comparison with Louhi Kasahara et al. [14], the only difference between those two methods being the active query scheme. For Cases 1 and 2, results in Figure 9a,b show that the proposed active query scheme achieves consistent, higher quality of weak supervision, resulting in a better performance in average and lower values of standard deviation.

The method of Louhi Kasahara et al. [15] employs both visual information and weak supervision, and it can be noted that this allows good performance in Cases 1 and 2, shown in Figure 9a,b. However, the results of Louhi Kasahara et al. [15] were lacking in consistency, shown by high values of standard deviation. The addition of the proposed active query scheme in our proposed method allows again to raise the average performance as well as increase the consistency of output.

Case 3, the results of which are reported in Figure 9c, seems to be a more difficult dataset. This is certainly due to the higher number of clusters since the two blocks were from different batches and defect types also vary. Without visual information, while the consistency of output was slightly increased by our proposed active query scheme compared to Louhi Kasahara et al. [14], the average performance slightly decreased. However, since the addition of visual information allows the proposed method to outperform the method of Louhi Kasahara et al. [15], it can be strongly suspected that the nature of errors was different when using the proposed active query scheme: those errors were easier to compensate for by the spatial estimator.

The proposed active query method allowed average performance increase for both Area 1 and Area 2, as shown in Figure 9d,e. Performance on the two datasets collected in field conditions are overall lower than in laboratory conditions. This is certainly due to the environmental noise in such outdoor settings coupled with the noise of the hammering module. The irregularity of the hammering grid pattern and the low number of samples compared to datasets obtained in laboratory conditions may have also have affected the performance, especially with regards to the spatial estimator.



**Figure 9.** Performance evaluation of several weakly supervised methods for the hammering test in laboratory conditions: (A) method of Louhi Kasahara et al. [14]; (B) RCA with the proposed active query; (C) method of Louhi Kasahara et al. [15]; and (D) proposed method. The averages of 20 runs with sets of 20 must-links each are reported. Error bars corresponds to one standard deviation.

## 8. Conclusions

An active weakly supervised method for defect detection in concrete structures using hammering, an acoustic inspection method, and visual information is proposed. By actively querying the human user, based on sample pair distance in audio feature space and physical space, the consistency of the quality of weak supervision was increased, resulting in an overall increase in performance. The proposed method can greatly help the human user provide adequate weak supervision and can be expected to be valuable for real inspections, where the number of samples to be considered can be colossal. Experiences in both laboratory conditions, using concrete test blocks, and field conditions, using a mock tunnel and automated hammering module, showed the effectiveness of the proposed method.

As future work, we would like to further explore the potential of active weak supervision. Some variants of RCA make use of dissimilar pairs [19] and an active query scheme for such pairs would certainly contribute to increase performance. We would also like to conduct additional experiments, at larger scale and with more varied defect configurations.

**Author Contributions:** Conceptualization, J.Y.L.K., A.Y. and H.A.; methodology, J.Y.L.K., A.Y. and H.A.; software, J.Y.L.K., A.Y. and H.A.; validation, J.Y.L.K., A.Y. and H.A.; formal analysis, J.Y.L.K., A.Y. and H.A.; investigation, J.Y.L.K., A.Y. and H.A.; resources, J.Y.L.K., A.Y. and H.A.; data curation, J.Y.L.K., A.Y. and H.A.; writing—original draft preparation, J.Y.L.K.; writing—review and editing, J.Y.L.K., A.Y. and H.A.; visualization, J.Y.L.K.; supervision, A.Y. and H.A.; project administration, A.Y. and H.A.; funding acquisition, J.Y.L.K., A.Y. and H.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by JSPS KAKENHI Grant Number 19J12391 and by the Japan Construction Information Center (JACIC) foundation.

**Acknowledgments:** The authors thank 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.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. BBC News. Japan Sasago Tunnel Collapse Killed Nine. Available online: <https://www.bbc.com/news/world-asia-20576492> (accessed on 11 April 2019).
2. BBC News. Italy Bridge Collapse: Genoa Death Toll Rises to 43. Available online: <https://www.bbc.com/news/world-europe-45241842> (accessed on 11 April 2019).
3. Cheng, Y.; Gao, F.; Hanif, A.; Lu, Z.; Li, Z. Development of a capacitive sensor for concrete structure health monitoring. *Constr. Build. Mater.* **2017**, *149*, 659–668. [CrossRef]
4. Giri, P.; Kharkovsky, S. Dual-laser integrated microwave imaging system for nondestructive testing of construction materials and structures. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 1329–1337. [CrossRef]
5. Li, W.; Kong, Q.; Ho, S.C.M.; Lim, I.; Mo, Y.; Song, G. Feasibility study of using smart aggregates as embedded acoustic emission sensors for health monitoring of concrete structures. *Smart Mater. Struct.* **2016**, *25*, 115031. [CrossRef]
6. Kim, J.; Gucunski, N.; Duong, T.H.; Dinh, K. Three-dimensional visualization and presentation of bridge deck condition based on multiple NDE data. *J. Infrastruct. Syst.* **2016**, *23*, B4016012. [CrossRef]
7. Henderson, M.E.; Dion, G.N.; Costley, R.D. Acoustic inspection of concrete bridge decks. In *Nondestructive Evaluation of Bridges and Highways III*; Society of Photo Optical: Newport Beach, CA, USA, 1999; pp. 219–227.
8. Zhang, G.; Harichandran, R.S.; Ramuhalli, P. An automatic impact-based delamination detection system for concrete bridge decks. *NDT & E Int.* **2012**, *45*, 120–127.
9. Fujii, H.; Yamashita, A.; Asama, H. Defect detection with estimation of material condition using ensemble learning for hammering test. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–20 May 2016; pp. 3847–3854.
10. Giri, P.; Mishra, S.; Clark, S.M.; Samali, B. Detection of gaps in concrete–metal composite structures based on the feature extraction method using piezoelectric transducers. *Sensors* **2019**, *19*, 1769. [CrossRef] [PubMed]
11. Ye, X.; Chen, X.; Lei, Y.; Fan, J.; Mei, L. An Integrated Machine Learning Algorithm for Separating the Long-Term Deflection Data of Prestressed Concrete Bridges. *Sensors* **2018**, *18*, 4070. [CrossRef] [PubMed]
12. Louhi Kasahara, J.Y.; Fujii, H.; Yamashita, A.; Asama, H. Clustering of Spatially Relevant Audio Data using Mel-Frequency Cepstrum for Diagnosis of Concrete Structure by Hammering Test. In Proceedings of the 2017 IEEE/SICE International Symposium on System Integration (SII), Taipei, Taiwan, 11–14 December 2017; pp. 787–792.
13. Louhi Kasahara, J.Y.; Fujii, H.; Yamashita, A.; Asama, H. Fuzzy Clustering of Spatially Relevant Acoustic Data for Defect Detection. *IEEE Rob. Autom. Lett.* **2018**, *3*, 2616–2623. [CrossRef]
14. Louhi Kasahara, J.Y.; Fujii, H.; Yamashita, A.; Asama, H. Weakly Supervised Approach to Defect Detection in Concrete Structures Using Hammering Test. In Proceedings of the International Conference on Consumer Electronics, Osaka, Japan, 15–18 October 2019; pp. 997–998.
15. Louhi Kasahara, J.Y.; Fujii, H.; Yamashita, A.; Asama, H. Complementarity of Sensors and Weak Supervision for Defect Detection in Concrete Structures. In Proceedings of the International Symposium on System Integration (SII), Honolulu, HI, USA, 12–15 January 2020; pp. 1–6.
16. Bar-Hillel, A.; Hertz, T.; Shental, N.; Weinshall, D. Learning distance functions using equivalence relations. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2003; pp. 11–18.
17. Ganchev, T.; Fakotakis, N.; Kokkinakis, G. Comparative evaluation of various MFCC implementations on the speaker verification task. In Proceedings of the SPECOM, Patras, Greece, 17–19 October 2005; pp. 191–194.
18. Müller, M. *Information Retrieval for Music and Motion*; Springer: Berlin/Heidelberg, Germany, 2007.
19. Yeung, D.Y.; Chang, H. Extending the relevant component analysis algorithm for metric learning using both positive and negative equivalence constraints. *Pattern Recognit.* **2006**, *39*, 1007–1010. [CrossRef]
20. Hoi, S.C.; Liu, W.; Lyu, M.R.; Ma, W.Y. Learning distance metrics with contextual constraints for image retrieval. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06), New York, NY, USA, 17–22 June 2006; pp. 2072–2078.
21. Hoi, S.C.; Liu, W.; Chang, S.F. Semi-supervised distance metric learning for collaborative image retrieval and clustering. *ACA Trans. Multimedia Comput. Commun. Appl. (TOMM)* **2010**, *6*, 18. [CrossRef]

22. Yang, J.; Xu, H. Metric learning based object recognition and retrieval. *Neurocomputing* **2016**, *190*, 70–81. [[CrossRef](#)]
23. Yang, C.; Tan, Y.; Bruzzone, L.; Lu, L.; Guan, R. Discriminative feature metric learning in the affinity propagation model for band selection in hyperspectral images. *Remote Sens.* **2017**, *9*, 782–798. [[CrossRef](#)]
24. Chuang, K.S.; Tzeng, H.L.; Chen, S.; Wu, J.; Chen, T.J. Fuzzy c-means clustering with spatial information for image segmentation. *Comput. Med. Imaging Graph.* **2006**, *30*, 9–15. [[CrossRef](#)] [[PubMed](#)]
25. Takahashi, Y.; Nakamura, S.; Ogawa, Y.; Satoh, T. Velocity Control Mechanism of the Under-actuated Hammering Robot for Gravity Compensation. In Proceedings of the 34th International Symposium on Automation and Robotics in Construction (ISARC), Taipei, Taiwan, 28 June–1 July 2017.
26. Nakamura, S.; Takahashi, Y.; Inoue, D.; Ueno, T. The Variable Guide Frame Vehicle for Tunnel Inspection. In Proceedings of the 34th International Symposium on Automation and Robotics in Construction (ISARC), Taipei, Taiwan, 28 June–1 July 2017.
27. Inoue, F.; Kwon, S.; Uchiyama, T.; Nakamura, S.; Yanagihara, Y. Shape Control of Variable Guide Frame for Tunnel Wall Inspection. In Proceedings of the 34th International Symposium on Automation and Robotics in Construction (ISARC), Taipei, Taiwan, 28 June–1 July 2017.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).