Weakly Supervised Acoustic Defect Detection in Concrete Structures Using Clustering-Based Augmentation

Jun Younes Louhi Kasahara, *Member, IEEE*, Hiromitsu Fujii, *Member, IEEE*, Atsushi Yamashita, *Senior Member, IEEE*, and Hajime Asama, *Fellow, IEEE*

***Abstract*—The automation of inspection methods for concrete structures is a pressing issue worldwide. Weakly supervised approaches, i.e., approaches based on supervision in other forms than traditional class labels, offer a unique mix of automation and human involvement that is highly effective for critical tasks such as inspection work. Generating weak supervision is less tedious than generating training data for supervised learning approaches. However, since it is less informative, high amounts of weak supervision are often needed. In practice, it is often the case that only scarce amounts of weak supervision are available. In this article, we propose a novel approach for weakly supervised acoustic defect detection in concrete structures that augment human-provided weak supervision. Experiments in both laboratory and field conditions showed that the proposed method allows for considerable performance gains for low amounts of weak supervision.**

***Index Terms*—Augmentation, clustering, defect detection, infrastructure inspection, weak supervision.**

Fig. 1. Hammering test being conducted in a tunnel.

I. Introduction

Modern societies are characterized by highly developed social infrastructures such as tunnels, highways, and bridges. Those are predominantly made out of concrete. Due to various factors ranging from aging to environmental conditions, defects appear over time in concrete structures. If left unchecked, those defects may lead to grave accidents, potentially leading to loss of lives and costly repairs. Recent unfortunate events such as the collapse of the ceiling panels of the Sasago tunnel in Japan [1] or the collapse of the Morandi bridge in Italy [2] have highlighted the need to regularly inspect concrete structures.

Among all the inspection methods, the hammering test, illustrated in Fig. 1, is one of the most popular methods. It consists of using a hammer to strike the surface of the concrete. The human inspector then uses the impact sound to look for defects hidden beneath the surface. The popularity of the hammering test in inspection sites is explained by its simplicity and effectiveness. However, it requires skilled human inspectors. Due to the manpower shortage and the growing number of structures in need of testing, the automation of the hammering test is highly sought after. More generally, the automation of inspection work is the focus of several researches in the field of mechatronics in construction.

There are previous studies that focused on the automation of the hammering test. The vast majority of them employed machine learning methods.

Supervised learning approaches represent most of the previous works. In [3], the sound of dragging chains across the surface of highways was analyzed using linear prediction coefficients. The works in [4] used ensemble learning with time-frequency analysis on hammering data and achieved high defect detection rates, with a classification of defects by their depth. In [5], a deep learning approach was implemented. Those approaches yielded remarkable results. However, due to their supervised nature, they rely heavily on the quantity and quality of the available training

Manuscript received January 13, 2021; revised March 18, 2021; accepted April 26, 2021. Date of publication May 4, 2021; date of current version December 15, 2021. Recommended by Technical Editor T. Linner and Senior Editor C. C. Cheah. This work was supported in part by JSPS KAKENHI under Grant Number JP21K17829, in part by the Satomi scholarship foundation, and in part by the Suzuki foundation.
(Corresponding author: Jun Younes Louhi Kasahara.)

Jun Younes Louhi Kasahara, Atsushi Yamashita, and Hajime Asama are with the Department of Precision Engineering, The University of Tokyo, Tokyo 113-8656, Japan (e-mail: louhi@robot.t.u-tokyo.ac.jp; yamashita@robot.t.u-tokyo.ac.jp; asama@robot.t.u-tokyo.ac.jp).

Hiromitsu Fujii is with the Department of Advanced Robotics, Chiba Institute of Technology, Chiba 275-0016, Japan (e-mail: hiromitsu.fujii@p.chibakoudai.jp).
Color versions of one or more figures in this article are available at https://doi.org/10.1109/TMECH.2021.3077496.

Digital Object Identifier 10.1109/TMECH.2021.3077496

data. Due to the high variability in composition, structure, and conditions of inspection, the gathering of appropriate training data for each structure can be expected to induce high logistical and manpower costs.

In order to bypass the issues related to training data, unsupervised learning approaches have been also proposed. In [6], cluster analysis with thresholding was proposed. Cluster analysis along with sensor fusion, where camera information was used along with acoustic information, was proposed in [7] and [8]. While good results were obtained, there was a need for additional information besides the acoustic information in order to achieve satisfactory levels of performance. This has the demerit of not being applicable to or requiring modification on already existing inspection apparatuses that are not equipped with the corresponding sensors.

Weakly supervised methods are methods based on supervision provided in other forms than class labels, which are traditional training data for supervised methods. Those other forms of supervision are generally easier to generate while having the drawback of being less informative than class labels. Weakly supervised methods can be thought of as being particularly suited for critical tasks such as inspection work: This is because they allow a subtle mix of human involvement and automation. In [9] and [10] was proposed an initial framework based on weakly supervised metric learning for defect detection in concrete structures, expanded in [11] with the addition of an active query scheme and [12] with the consideration of multiple humans providing the weak supervision. This framework was based on queries to human users on the similarity of selected sample pairs. However, the performance was conditioned by the number of allowed queries: If not enough queries were allowed, the defect detection performance would drop significantly. Since queries involve human participation, it is desirable that the number of queries is kept at a minimum.

The objective of this article is therefore to propose a novel weak supervision augmentation method for acoustic defect detection in concrete structures. This allows the generation of *artificial* weak supervision to complement the weak supervision provided by humans and, therefore, high defect detection performance in cases of low amounts of available weak supervision.

II. Method

A. Concept and Overview

Concretely, weak supervision refers to pairs of samples that a human user has designated through queries as being similar, referred to as *must-links*, or dissimilar, referred to as *cannot-links*. Both are also referred to as *constraints* in the literature. Previous works in the field usually consider the availability of large amounts of weak supervision in their experiments, which is difficult to achieve in practice since queries are answered by humans.

There has been little attention given to the augmentation of weak supervision in the literature. The studies in [13] and [14] used an ensemble learning framework to find pairs often attributed to the same cluster and generate artificial must-links. This process is, however, purely unsupervised and is

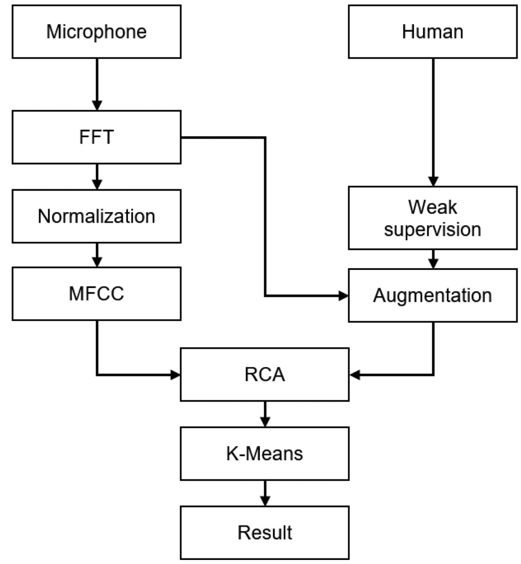


Fig. 2. Overview of the proposed method.

uncorrelated with the initially provided weak supervision, i.e., it does not take advantage of the human-provided knowledge.

Our proposed approach aims to produce new knowledge by taking advantage of both unsupervised clustering and human-provided weak supervision: Human-provided weak supervision, corresponding to the constraints originally provided by a human, is supplemented in our proposed method by artificial weak supervision, i.e., weak supervision derived from human-provided weak supervision by clustering. It consists of using an initial partitioning of the dataset with a setting of high number of clusters in order to obtain small clusters of high purity. Each human-provided weak supervision pair is then used to generate several new pairs based on the initial partitioning.

Our contributions are as follows:

1) A novel concept aiming to augment weak supervision for defect detection in concrete structures, based on both unsupervised clustering and human-provided weak supervision, is introduced. This enables considerable performance increase for cases of low amounts of weak supervision.
2) Testing of our proposed approach in both laboratory and field conditions has been conducted.

Weakly supervised learning for low amounts of weak supervision, a realistic and most often encountered situation in practice, rarely considered in the literature, is the main focus of the present article. An overview of our proposed method is shown in Fig. 2. Hammering sound samples are first converted into Fourier spectrum, normalized, and finally converted into Mel Frequency Cepstrum Coefficients (MFCCs). This serves as the initial feature space. Human-provided weak supervision is augmented using the hammering samples dataset in their Fourier space. Then, weakly supervised feature space transformation using the augmented weak supervision is conducted using Relevant Component Analysis (RCA) on the hammering sample data in MFCC space. Finally, the discrimination between defect and nondefect hammering samples is done using K-means.

B. Preprocessing and Initial Feature Space

Hammering samples are initially short audio segments of about 20 ms corresponding to the initial response to the hammer strike. Using the Fast Fourier Transform (FFT), hammering samples are, therefore, first converted from time-series data into Fourier spectrum in a similar fashion as in [4]. Normalization to zero-mean and unit-variance is conducted following the procedure described in [11]. Then, transformation to MFCC is conducted. MFCCs are feature vectors originally designed to mimic the human auditory perception and popularly used in audio recognition tasks [15]. Our previous works showed the effectiveness of MFCC at separating defect and nondefect hammering samples [7]. Therefore, MFCCs are here used as the initial feature space.

C. Weak Supervision Augmentation

1) Fine Partitioning: Difficult datasets, i.e., when the appropriate similarity metric is not known, are troublesome to cluster with the given number of clusters K. However, if clustering is conducted with a number of clusters well in excess, local structures of the data, which are much easier to discern, will be matched by most of the created clusters. More concretely, unless the choice of default metric is extremely bad, two very similar samples should still only have a small distance separating them: The shortcomings of an inappropriate metric manifest themselves over medium and long distances in the resulting feature space.

Therefore, the first step of our proposed method is to obtain a partitioning of the dataset with a number of clusters K_c far exceeding the number of clusters of the dataset K. This fine partitioning does not aim to cluster the dataset correctly between classes but to generate clusters that do not mix samples of different classes. From an image processing point of view, this would be similar to generating a high amount of segmentation borders in the hope of obtaining all the correct borders among the meaningless ones.

This fine partitioning should be conducted with $K_c \gg K$, i.e., a number of clusters far superior to the initial number of classes. This can be regulated through the introduction of a new arbitrary parameter γ in the setting of K_c such as in (1), with N_s being the number of samples of the considered dataset and [.] being the round function. Under the assumption that the unsupervised clustering process used to generate this fine partitioning tends to produce partitions of equal sizes, γ can be seen as regulating the amount of new artificial weak supervision generated on top of the human-provided weak supervision

$$K_c = \left[\frac{N_s}{\gamma}\right]. \quad (1)$$

2) Possible Pairs and Types of Relationships: Once a fine partitioning of the dataset has been obtained, human-provided weak supervision can be expanded, i.e., constraints between samples can be elevated to constraints between clusters.

Given a fine partitioning $P = \{C_k\}_{k\in[1...K_c]}$ of the dataset, with C_k being a cluster, and a constraint $\{\mathbf{a}, \mathbf{b}\}$ such as $\mathbf{a} \in C_a$ and $\mathbf{b} \in C_b$, with $a, b \in [1...K_c]$, three types of relationships between samples can be distinguished, as illustrated in Fig. 3.

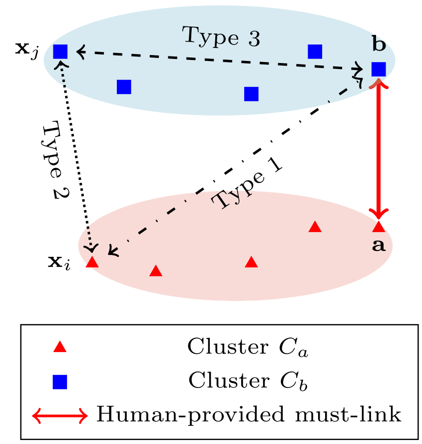


Fig. 3. Illustration of the different types of relationships that can be generated based on an user-provided must-link and given an initial clustering of the dataset.

1) Type 1 relationships, between a sample involved in a constraint and a sample belonging to the cluster of the other sample of the pair, e.g., between $\{\mathbf{b}, \mathbf{x}_i\}$ with $\mathbf{x}_i \in C_a$.
2) Type 2 relationships, between samples that are not part of the constraint but belonging to different clusters across the constraint, e.g., between $\{\mathbf{x}_i, \mathbf{x}_j\}$ with $\mathbf{x}_i \in C_a$ and $\mathbf{x}_j \in C_b$. The differences with Type 1 relationships being that both samples are not comprised in the constraint set provided by the user.
3) Type 3 relationships, inside each cluster of the partition, $\{\mathbf{a}, \mathbf{x}_i\}$ or $\{\mathbf{x}_i, \mathbf{x}_{i'}\}$ with $\mathbf{x}_i \in C_a$ and $\mathbf{x}_{i'} \in C_a$ for example.

Type 3 relationships only concern must-link constraints and can be argued not to be relevant in the considered case. Indeed, those do not actually contain any additional information provided by the user through the constraint $\{\mathbf{a}, \mathbf{b}\}$; they were obtained in a purely unsupervised fashion. Feeding back such kind of constraints would potentially bring the final behavior closer to the behavior of the process used to generate the fine partition P. This could be mitigated to a certain extent through an ensemble framework as in [13] and [14].

With the goal of expanding human-provided constraints, Type 1 and Type 2 relationships are of interest since those mix information provided by humans and information obtained through unsupervised means: the intercluster liaison is human-provided and the intracluster liaisons were found by unsupervised clustering.

Between Type 1 and Type 2, the latter is most likely to bring new contributions to the metric learning process. Indeed, those involve most of the time two samples with both of them not being involved in a previous constraint, allowing to provide information on totally new areas of the feature space.

3) Informativeness of Type 2 Relationships: Given a constraint and the corresponding cluster of each of the pairs' elements, there may be a tradeoff between information gain and error risk. Type 2 relationships in close proximity of user-provided

Algorithm 1: Proposed Weak Supervision Augmentation Scheme for Must-Links.

Input: dataset $D = \{\mathbf{x}_i\}_{i\in[1\ldots N_s]}$, partitioning $P = \{C_k\}_{k\in[1\ldots K_c]}$ of D, must-links $\mathcal{M}$
Output: augmented must-links $\mathcal{M}^*$
for each $\{\mathbf{a},\mathbf{b}\} \in \mathcal{M}$ **do**
Add $\{\mathbf{a},\mathbf{b}\}$ to $\mathcal{M}^*$
Find the corresponding clusters of P so that $\mathbf{a} \in C_a = \{\mathbf{x}_j^a\}_{j=[1\ldots|C_a|]}$ and $\mathbf{b} \in C_b = \{\mathbf{x}_j^b\}_{j=[1\ldots|C_b|]}$
Order each element of each cluster in decreasing distance from samples involved in must-link
for $l = 1$ **to** $\min(|C_a|, |C_b|)$ **do**
  if $\{\mathbf{x}_l^a, \mathbf{x}_l^b\} \notin \mathcal{M}^*$ **then**
    Add $\{\mathbf{x}_l^a, \mathbf{x}_l^b\}$ to $\mathcal{M}^*$
  end if
end for
end for

constraint have a high probability of being correct, but they are also more likely to only be able to provide redundant information. The further the pairs of potential new Type 2 relations are from the user-provided constraint, the more potential they hold to bring new information useful to the RCA transformation and thus achieve an appropriate feature space shape. However, they are also more likely to be erroneous.

4) Proposed Constraint Augmentation Scheme: For each human-provided constraint, the corresponding clusters of each of the two elements composing the pair obtained by fine partitioning are considered. In each of those clusters, one element is involved in the human-provided constraint. Other elements are ordered in decreasing distance using the default metric from this element involved in the user-provided constraint. Finally, new pairs are formed between the ordered elements across clusters until all the elements in a cluster have been used up.

While this scheme does not take into account the previously discussed tradeoff in Section II-C3, it has the merit of obtaining Type 2 relations in a manner that is simple and parallel to the human-provided constraint to a certain degree. Furthermore, the risk of obtaining erroneous constraints at far distances can be limited by the setting of the fine partitioning parameter γ.

The proposed constraint augmentation process in the case of must-links is described in Algorithm 1 and illustrated in Fig. 4.

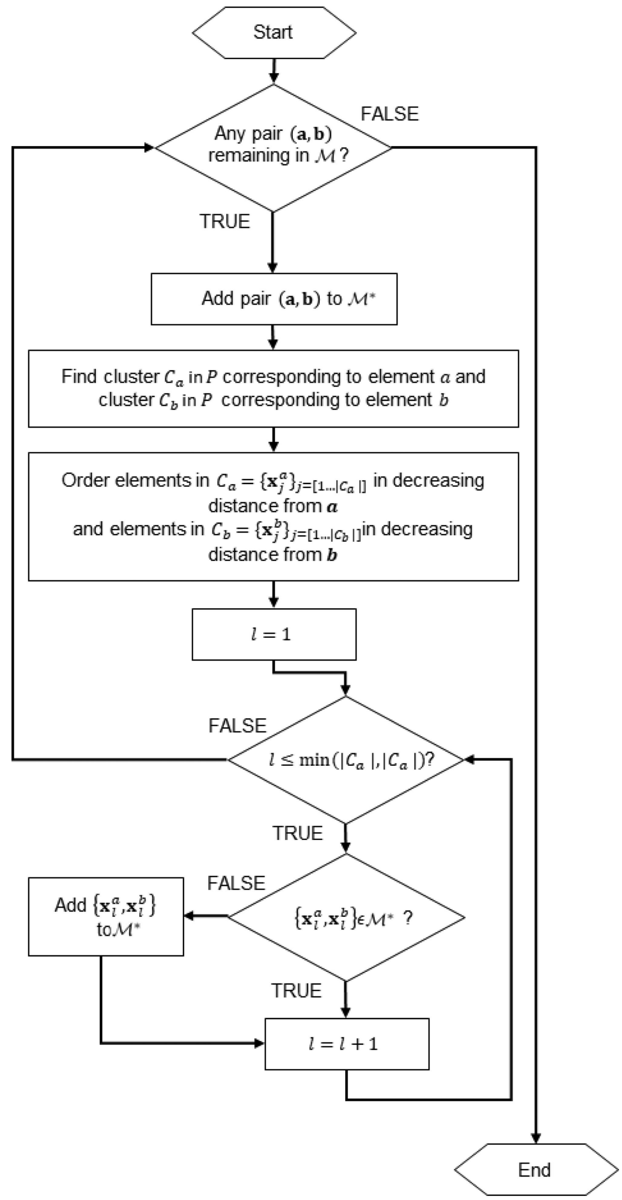


Fig. 4. Flow diagram of the proposed weak supervision augmentation scheme for must-links.

D. Weakly Supervised Feature Space Transformation

The work in [16], based on the initial work published in [17], proposed a procedure called RCA, a linear transformation of the feature space prior to clustering, similar to a whitening process. RCA is conducted based on must-link constraints. More precisely, RCA is conducted on chunklets which are samples that are deduced to belong to the same cluster from the provided constraints using the transitive property of must-links.

Given N_{chunklet} chunklets $\{\mathcal{M}_l\}_{l\in[1\ldots N_{\text{chunklet}}]}$, 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 (2), with N_{total} being the total number of elements contained in the chunklets. Here, $\hat{\mathbf{m}}_l$ is computed as the average vector of elements in $\mathcal{M}_l$

$$\hat{\mathbf{C}} = \frac{1}{N_{\text{total}}} \sum_{j=1}^{N_{\text{chunklet}}} \sum_{\mathbf{x}_i \in \mathcal{M}_l} (\mathbf{x}_i - \hat{\mathbf{m}}_l)(\mathbf{x}_i - \hat{\mathbf{m}}_l)^{\mathrm{T}}. \quad (2)$$

3) Compute the whitening transformation associated with this covariance matrix and apply it to the dataset: As in (3), the inverse square root of the covariance matrix $\hat{\mathbf{C}}$ is applied to each sample $\mathbf{x}$ of the initial dataset to obtain a new sample $\mathbf{x}_{\text{new}}$.

$$\mathbf{x}_{\text{new}} = \hat{\mathbf{C}}^{-1/2}\mathbf{x}. \quad (3)$$

Most recent studies with RCA involve the inclusion of cannot-links such as in [18] and [19] and kernelization in order to allow handling of nonlinear metrics, such as in [20]. RCA-like approaches have also seen success in several applications, notably in image pattern recognition, as in [21] and [22].

As it can be observed in the previously presented steps of RCA, inversion of the covariance matrix is required and this may be troublesome, especially when dealing with high-dimensional data, as pointed out in [23]. $\hat{\mathbf{C}}$ is a $d \times d$ square matrix and, therefore, is nonsingular only if its rank is equal to d.

The work in [18] provided an alternative formulation for $\hat{\mathbf{C}}$, as in (4), which lends an easier form for $\hat{\mathbf{C}}$ to evaluate its rank, with the $|.|$ operator denoting the cardinality of a set

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

Starting from (5) and given that if two matrices $\mathbf{A}$ and $\mathbf{B}$ are conformable, $\mathrm{rank}(\mathbf{A}+\mathbf{B}) \leq \mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B})$, (6) can be obtained

$$\mathrm{rank}((\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^{\mathrm{T}}) \leq 1, \tag{5}$$

$$\mathrm{rank}(\hat{\mathbf{C}}) \leq |\mathcal{M}|. \tag{6}$$

Therefore, the rank of the covariance matrix $\hat{\mathbf{C}}$ in RCA is inferior or equal to the number of provided must-links $|\mathcal{M}|$. If $|\mathcal{M}| \leq d$, then $\hat{\mathbf{C}}$ cannot be inversible. This results in the need for the number of must-links $|\mathcal{M}|$ to be at least equal to the dimensionality of the data d in order to be able to conduct RCA properly. By augmenting weak supervision, our proposed method allows to mitigate this requirement for RCA.

III. Experiments

Experiments were conducted in both laboratory and field conditions.

For experiments in laboratory conditions, concrete test blocks containing simulated defects were used and two cases were considered using the setup illustrated in Fig. 5(a).

1) *Case 1:* a single block containing a single defect, shown in Fig. 5(b). Four hundred and sixty two hammering samples were collected: 272 nondefects and 190 defects. The defect was a delamination at an angle of 30° from the surface running for a length of 200 mm on the surface.
2) *Case 2:* two blocks each containing a single defect, shown in Fig. 5(c). Two hundred and seventy hammering samples were collected: 155 nondefects and 115 defects. Both defects were delaminations at an angle of 15° from the surface.

A hammer commonly used by human inspectors in the field was used to conduct hammering (KTC UDHT-2, with head diameter of 16 mm, length of 380 mm, and weight of 160 g). To record the acoustic data, a Behringer ECM8000 microphone was used along with a Roland UA-25EX soundboard. Hammering samples were recorded at the standard 44.1 kHz. Fourier spectrums of length 1024 were obtained using FFT and MFCC were computed with 10 coefficients.

For experiments in field conditions, a mock tunnel was used. This tunnel, shown in Fig. 6(a), was built according to the same standards as current in-service tunnels and contains defects that occurred naturally. This provides us with a setting identical to actual inspection sites without the logistical and legal issues. Furthermore, hammering was conducted here using the automated hammering apparatus developed in [24]–[26] and shown in Fig. 6(b). The selected area for inspection is shown in Fig. 6(c) and presents a delamination.

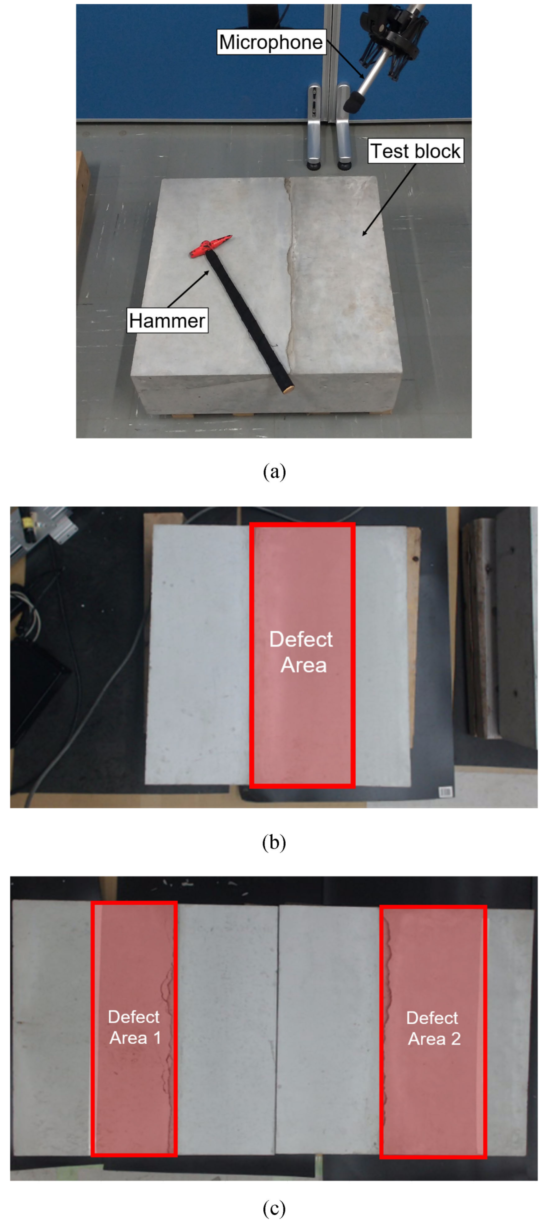


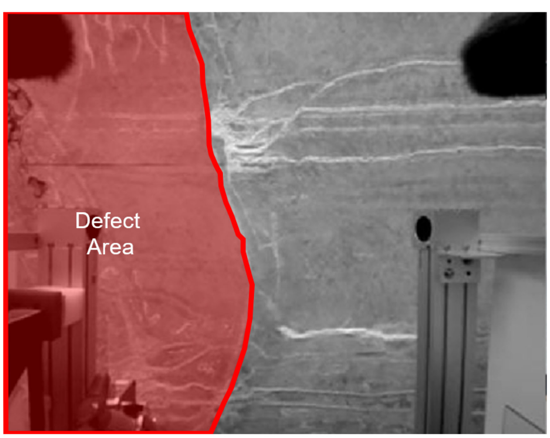
Fig. 5. Illustration of the different types of relationships that can be generated based on an user-provided must-link and given an initial clustering of the dataset. (a) Experimental setup in laboratory conditions. (b) Case 1: single delamination. The red overlay shows the defect area. (c) Case 2: dual delaminations. The red overlays show the defect areas.

The Rand index [15], a common measure of performance for clustering methods, was used to measure performance. The Rand index ranges between 0 and 1 and the closer the value is to 1, the better the method is. The following methods were compared in the experiments using the setting $K = 2$.

1) The method of [9] consisting of RCA over MFCC.

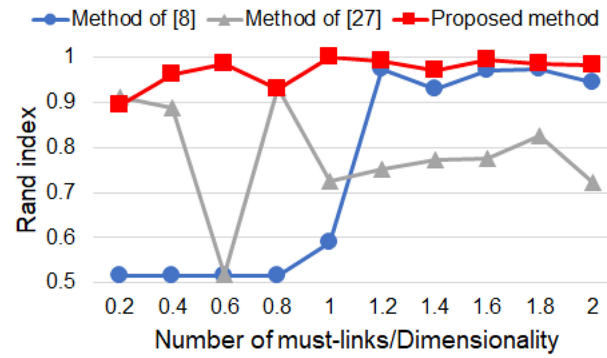
(a)

(b)

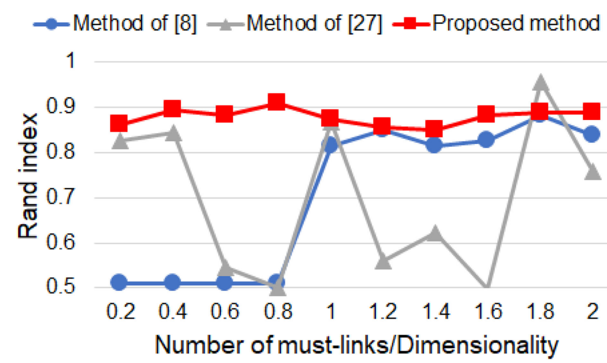


(c)

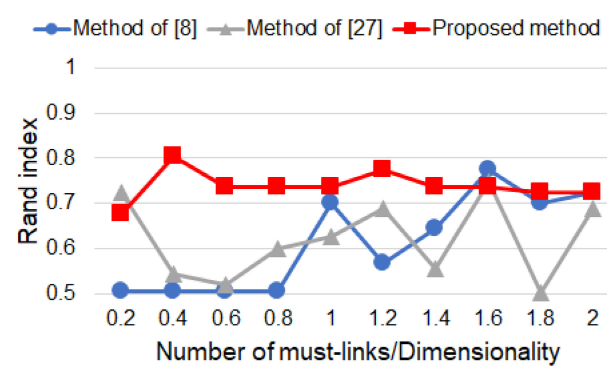
Fig. 6. Picture of the considered cases in laboratory conditions. Red areas indicate defect areas. (a) Experimental setup in field conditions. (b) Closeup view of the hammering module developed in [24]–[26]. (c) Tested area in the mock tunnel. The red overlay shows the defect area.



(a)



(b)



(c)

Fig. 7. Defect detection performance for various amounts of human-provided weak supervision. Best outputs out of 20 iterations are reported. (a) Performance in Case 1 for various amounts of human-provided weak supervision. (b) Performance in Case 2 for various amounts of human-provided weak supervision. (c) Performance in field experiment for various amounts of human-provided weak supervision.

2) The method of [27] consisting of RCA with a dimensionality reduction step.
3) The proposed method consisting of RCA with augmented weak supervision. The setting $\gamma = 3$ was used for all datasets.

IV. Results and Discussions

In Fig. 7 are reported the best output out of 20 iterations for each amount of weak supervision. The amount of provided weak supervision is shown as a ratio of the provided number of must-links over the dimensionality, and the pairs of samples over which to provide weak supervision were selected at random. In Tables I–III is reported the correctness of the augmented weak supervision using the proposed method for each experiment for various amounts of weak supervision.

Comparing with the method of [9], the proposed method effectively allows RCA to be competitive with significantly fewer amounts of constraints. Across all the three considered

TABLE I
NUMBER OF HUMAN-PROVIDED MUST-LINKS, AVERAGE TOTAL NUMBER OF MUST-LINKS AFTER USING TRANSITIVE CLOSURE TO OBTAIN NEW CONSTRAINTS (CHUNKLETS) AND AVERAGE TOTAL NUMBER OF CONSTRAINTS AFTER USING THE PROPOSED METHOD, WITH THE CORRESPONDING AVERAGE CORRECTNESS IN CASE 1

$\|\mathcal{M}\|$	**Chunklets**	**Augmentation**	
		$\|\mathcal{M}^*\|$	Correctness
2	2	7.7	100.0%
4	4	15.1	100.0%
6	6	23.0	100.0%
8	9	31.0	99.6%
10	10	43.8	98.4%
12	15	45.8	100.0%
14	16	54.5	100.0%
16	17	59.7	100.0%
18	19	72.6	100.0%
20	23	71.5	99.6%

TABLE II
NUMBER OF HUMAN-PROVIDED MUST-LINKS, AVERAGE TOTAL NUMBER OF MUST-LINKS AFTER USING TRANSITIVE CLOSURE TO OBTAIN NEW CONSTRAINTS (CHUNKLETS) AND AVERAGE TOTAL NUMBER OF CONSTRAINTS AFTER USING THE PROPOSED METHOD, WITH THE CORRESPONDING AVERAGE CORRECTNESS IN CASE 2

$\|\mathcal{M}\|$	**Chunklets**	**Augmentation**	
		$\|\mathcal{M}^*\|$	Correctness
2	2	9.1	100.0%
4	4	18.4	100.0%
6	6	22.3	95.5%
8	9	31.8	98.1%
10	10	46.1	96.8%
12	15	51.0	97.6%
14	16	61.5	94.5%
16	16	66.6	94.7%
18	20	71.1	96.0%
20	22	82.9	94.5%

TABLE III
NUMBER OF HUMAN-PROVIDED MUST-LINKS, AVERAGE TOTAL NUMBER OF MUST-LINKS AFTER USING TRANSITIVE CLOSURE TO OBTAIN NEW CONSTRAINTS (CHUNKLETS) AND AVERAGE TOTAL NUMBER OF CONSTRAINTS AFTER USING THE PROPOSED METHOD, WITH THE CORRESPONDING AVERAGE CORRECTNESS IN FIELD EXPERIMENT

$\|\mathcal{M}\|$	**Chunklets**	**Augmentation**	
		$\|\mathcal{M}^*\|$	Correctness
2	3	9.1	60.2%
4	4	13.1	80.5%
6	6	25.3	74.0%
8	8	33.7	79.5%
10	13	38.1	82.1%
12	14	49.7	87.7%
14	19	59.7	87.6%
16	19	60.4	81.8%
18	24	75.9	82.1%
20	30	83.2	92.1%

hammering datasets, RCA required around ten must-links to start achieving decent outputs. The proposed method enabled to reach such levels of performance at half or fewer amounts of weak supervision.

In Case 1, the method of [9], consisting of K-means over the feature space transformed by RCA, achieves a high Rand index value of around 0.95 for more than 10 must-links. For less than 10 must-links, i.e., a ratio of must-links/dimensionality less than 1, RCA created feature spaces where K-means was only able to perform as well as a random classifier. Those are cases of failure: As explained earlier, RCA-type approaches require a minimum amount of constraints; otherwise, singularity of the covariance matrix does not allow to obtain valid features. In fact, a little more is needed, as competitive performance is met at around a must-links/dimensionality ratio of 1.2. The proposed augmentation scheme allows RCA here to be applicable with significantly fewer amounts of weak supervision. According to Table I, the proposed method allowed reaching the threshold of ten must-links from only two constraints provided by humans. Past this threshold, the proposed method also allows the method of [9] to have higher performance.

RCA with dimensionality reduction [27] showed an unstable behavior with the considered experimental settings. This could be explained by two factors. First, RCA with dimensionality reduction involves principal component analysis, which is variance-based and may be unsuitable for the considered task. Second, RCA with dimensionality reduction also includes a variant of Fisher linear discriminant on top of the transformation induced by RCA. This could be applying too much bias based on the provided constraints, i.e., the process is very sensitive to the quality of provided constraints when the amount of weak supervision is low.

In Case 2, all methods returned lower values of Rand index than for Case 1, showing that Case 2 was a more difficult dataset than Case 1 due to the presence of two defects. The method of [9] exhibited about the same behavior as in Case 1. According to Table II, while the amount of generated additional constraints using the proposed method was about the same, the correctness of those new constraints was lower than that for Case 1. This is explained by the increase in difficulty, i.e., even with a fine partitioning, some local structures of the data have not been recognized. However, a small percentage of erroneous constraints only marginally affects RCA [28]. Hence, the proposed constraint augmentation scheme allows having a better performance with low amounts of weak supervision. Results on RCA with dimensionality reduction were also very similar to the ones obtained for Case 1.

The overall performances of the considered methods in the field experiments were lower than those in the laboratory experiments. This is due to the environmental noise since the mock tunnel was located outside as well as the motor noise of the hammering apparatus. The difficulty of this dataset is translated by the method of [9] having issues exhibiting a stable performance at over a Rand index value of 0.7 until a ratio of must-links/dimensionality of 1.6. On the other hand, our proposed method achieved that threshold of performance with much less weak supervision. The method of [27] once again showed unstable performance.

Artificially generated content is generally less reliable than original content, i.e., artificial weak supervision can reasonably be expected to have lower correctness than human weak supervision. This is offset to a certain extent in our proposed method by the use of RCA. RCA boasts high robustness to errors in weak supervision since the major axis of the transformations are computed based on the majority of the provided pairwise constraints, additional constraints mainly serving as fine-tuning. This was also shown experimentally on hammering data in our previous work [28], where, even for significant amounts of erroneous weak supervision (up to 25% error rate), high clustering performance was obtained.

In practice, we would recommend keeping the parameter for the augmentation γ as low as possible. This is because this parameter controls how detailed the fine partition is for the augmentation: The finer this partition is, fewer the errors that are expected among artificial weak supervision. Furthermore, the main advantage of our proposed method is for cases of low amounts of human weak supervision, which leads to a catastrophic drop in performance for previous works as shown in our experimental section. Once our proposed method of augmentation has allowed for sufficient artificial weak supervision to be generated, the performance gain compared to only using human weak supervision is lesser.

In the construction industry, suffering worldwide from issues related to manpower shortage, robotics systems for automation of construction bear high expectations. Those are indicated by the multitude of researches in the field [29], [30]. Construction work is not limited to the actual construction: Inspection work is also a major component and targets both newly built structures, to ensure conformity to specifications, and pre-existing structures, to monitor deterioration levels, and to assess the need for repairs/maintenance work. Inspection work is still predominantly conducted by humans. The application of our proposed method along with other robotic systems focused on the automation of construction would allow automation for the whole lifespan of concrete structures.

As in our field experiments, our proposed method can be used with robotic systems such as the hammering apparatus developed in [24]–[26]. This would allow the automation of the full inspection process. Furthermore, such systems could be integrated into an Internet-based platform for information sharing to allow the analysis and storage of inspection data across multiple concrete structures.

V. Conclusion

In this article, we have the following:

1) A novel approach to augment weak supervision for acoustic defect detection in concrete structures was proposed. Artificial weak supervision was generated based on the human-provided weak supervision and used to increase the overall amount of supervision. This is especially beneficial in the conditions of low amounts of available weak supervision, which is often the case in practice. This can be expected to greatly benefit the spread of mechatronics systems for automated inspection work in construction.
2) Experiments in both laboratory and field conditions showed that the proposed method significantly improves performance when the amount of weak supervision is low. Furthermore, the proposed method is also beneficial for cases where the amount of weak supervision is sufficient as well.

In the future, refinements in the constraints augmentation scheme, including an evaluation of uncertainty, would certainly be beneficial and increase overall performance by allowing to filter out some of the erroneous constraints generated by our proposed method. Extending this framework to include active constraint selection, i.e., actively querying the user, could also be interesting and is a topic that has attracted much attention in the field of weakly supervised clustering. Additionally, integration into an unified information system for the entire inspection and maintenance process would be desirable.

References

[1] BBC News, "Japan Sasago tunnel collapse killed nine," 2012. Accessed: Nov. 4, 2019. [Online]. Available: https://www.bbc.com/news/world-asia-20576492
[2] BBC News, "Italy bridge collapse: Genoa death toll rises to 43," 2018. Accessed: Nov. 4, 2019. [Online]. Available: https://www.bbc.com/news/world-europe-45241842
[3] M. E. Henderson, G. N. Dion, and R. D. Costley, "Acoustic inspection of concrete bridge decks," in *Proc. Nondestruct. Eval. Techn. Aging Infrastructures Manuf. Int. Soc. Opt. Photon.*, 1999, pp. 219–227.
[4] 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.
[5] T. Fukumura, H. Aratame, A. Ito, M. Koike, K. Hibino, and Y. Kawamura, "A study on hammering test using deep learning," in *Proc. IEEE 15th Int. Conf. Syst. Syst. Eng.*, 2020, pp. 261–264.
[6] J. Y. Louhi Kasahara, H. Fujii, A. Yamashita, and H. Asama, "Unsupervised learning approach to detection of void-type defects in concrete structure using hammering and clustering," in *Proc. Int. Conf. Adv. Mechatron., JSME*, 2015, pp. 319–320.
[7] 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. Integration*, 2017, pp. 787–792.
[8] J. Y. Louhi Kasahara, H. Fujii, A. Yamashita, and H. Asama, "Fuzzy clustering of spatially relevant acoustic data for defect detection," *IEEE. Robot. Autom. Lett.*, vol. 3, no. 3, pp. 2616–2623, Jul. 2018.
[9] J. Y. Louhi Kasahara, H. Fujii, A. Yamashita, and H. Asama, "Weakly supervised approach to defect detection in concrete structures using hammering test," in *Proc. Int. Conf. Consum. Electron.*, 2019, pp. 997–998.
[10] J. Y. L. Kasahara, A. Yamashita, and H. Asama, "Weakly supervised defect detection using acoustic data based on positive and negative constraints," in *Proc. IAARC 37th Int. Symp. Autom. Robot. Construction*, 2020, pp. 1331–1336.
[11] J. Y. Louhi Kasahara, A. Yamashita, and H. Asama, "Acoustic inspection of concrete structures using active weak supervision and visual information," *Sensors*, vol. 20, no. 3, 2020, Art. no. 629.
[12] J. Y. L. Kasahara, A. Yamashita, and H. Asama, "Multiple oracle consensus for weakly supervised defect detection in concrete structures using audio data," in *Advanced Robotics*. New York, NY, USA: Taylor & Francis, 2020.
[13] D. Greene and P. Cunningham, "Constraint selection by committee: An ensemble approach to identifying informative constraints for semi-supervised clustering," in *Proc. Eur. Conf. Mach. Learn.*, 2007, pp. 140–151.
[14] I. Diaz-Valenzuela, V. Loia, M. J. Martin-Bautista, S. Senatore, and M. A. Vila, "Automatic constraints generation for semisupervised clustering: Experiences with documents classification," *Soft Comput.*, vol. 20, no. 6, pp. 2329–2339, 2016.
[15] M. Müller, *Information Retrieval for Music and Motion*. Berlin, Germany: Springer, 2007.
[16] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning distance functions using equivalence relations," in *Proc. Int. Conf. Mach. Learn.*, 2003, pp. 11–18.
[17] N. Shental, T. Hertz, D. Weinshall, and M. Pavel, "Adjustment learning and relevant component analysis," in *Proc. Eur. Conf. Comput. Vis.*, Berlin, Germany: Springer, 2002, pp. 776–790.
[18] D.-Y. Yeung and H. Chang, "Extending the relevant component analysis algorithm for metric learning using both positive and negative equivalence constraints," *Pattern Recognit.*, vol. 39, no. 5, pp. 1007–1010, 2006.
[19] S. C. Hoi, W. Liu, M. R. Lyu, and W.-Y. Ma, "Learning distance metrics with contextual constraints for image retrieval," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Rattern Recognit.*, vol. 2, 2006, pp. 2072–2078.
[20] S. C. Hoi, W. Liu, and S.-F. Chang, "Semi-supervised distance metric learning for collaborative image retrieval and clustering," *Trans. Multimedia Comput., Commun., Appl.*, vol. 6, no. 3, 2010, pp. 1–26.

[21] J. Yang and H. Xu, "Metric learning based object recognition and retrieval," *Neurocomputing*, vol. 190, pp. 70–81, 2016.

[22] C. Yang, Y. Tan, L. Bruzzone, L. Lu, and R. Guan, "Discriminative feature metric learning in the affinity propagation model for band selection in hyperspectral images," *Remote Sensing*, vol. 9, no. 8, pp. 782–798, 2017.

[23] S. Xiang, F. Nie, and C. Zhang, "Learning a Mahalanobis distance metric for data clustering and classification," *Pattern Recognit.*, vol. 41, no. 12, pp. 3600–3612, 2008.

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

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

[26] S. Nakamura *et al.*, "Inspection test of a tunnel with an inspection vehicle for tunnel lining concrete," *J. Robotics Mechatron.*, vol. 31, no. 6, pp. 762–771, 2019.

[27] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning a Mahalanobis metric from equivalence constraints," *J. Mach. Learn. Res.*, vol. 6, pp. 937–965, Jun. 2005.

[28] J. Y. Louhi Kasahara, H. Fujii, A. Yamashita, and H. Asama, "Complementarity of sensors and weak supervision for defect detection in concrete structures," in *Proc. Int. Symp. Syst. Integration*, 2020, pp. 1–6.

[29] T. Bock, "The future of construction automation: Technological disruption and the upcoming ubiquity of robotics," *Autom. Construction*, vol. 59, pp. 113–121, 2015.

[30] H. Ardiny, S. Witwicki, and F. Mondada, "Construction automation with autonomous mobile robots: A review," in *Proc. 3rd RSI Int. Conf. Robot. Mechatron.*, 2015, pp. 418–424.

Jun Younes Louhi Kasahara (Member, IEEE) received the B.E. and Diplôme d'Ingénieur degrees from the Ecole Centrale de Lyon, Écully, France, in 2013 and 2016, respectively, and the Ph.D. degree in engineering from The University of Tokyo, Tokyo, Japan, in 2019.

He was a specially appointed Researcher with The University of Tokyo from 2019 to 2020. Since 2020, he has been a Project Assistant Professor with The University of Tokyo. His main research interests include machine learning, algorithms, signal processing, and automation.

Dr. Kasahara is also a Member of the JSPE.

Hiromitsu Fujii (Member, IEEE) received the B.E. and M.E. degrees in engineering and the Ph.D. degree in precision engineering from The University of Tokyo, Tokyo, Japan, in 2005, 2007, and 2016, respectively.

He worked as an image processing Researcher and a Software Engineer with Sony Corporation Ltd. Tokyo, Japan, from 2007 to 2013, and he was a JSPS Research Fellow from 2014 to 2016. From 2016 to 2017 and from 2017 to 2018, he was an Assistant Professor and a Lecturer with The University of Tokyo, respectively. He is currently an Associate Professor with the Department of Advanced Robotics, Chiba Institute of Technology, Chiba, Japan. His research interests include signal processing for robotics systems, such as multisensor integration for unmanned tele-operation systems and pattern recognition for the automated diagnoses using robots.

Dr. Fujii is also a Member of JSME, SICE, and RSJ.

Atsushi Yamashita (Senior Member, IEEE) received B.E., M.E., and Ph.D. degrees in Engineering from the Faculty of Engineering, The University of Tokyo, Tokyo, Japan, in 1996, 1998, and 2001, respectively.

He was an Assistant Professor with Shizuoka University, Shizuoka, Japan, from 2001 to 2008. From 2006 to 2007, he was a Visiting Associate with the California Institute of Technology, Pasadena, CA, USA. From 2008 to 2011, he was an Associate Professor with Shizuoka University. Since 2011, he has been an Associate Professor with The University of Tokyo. His research interests include robot vision, image processing, and intelligent robot system.

Dr. Yamashita is also a Member of ACM, JSPE, RSJ, IEICE, JSAE, JSME, IEEJ, IPSJ, ITE, SICE, and Society for Serviceology.

Hajime Asama (Fellow, IEEE) received the M.S. and Dr.Eng. degrees from The University of Tokyo, Tokyo, Japan (UTokyo) in 1984 and 1989, respectively, both in engineering.

He was with RIKEN, Wako, Japan, from 1986 to 2002, became a Professor of Research into Artifacts, Center for Engineering (RACE), UTokyo in 2002, has been a Professor with School of Engineering since 2009, and has been the Director of RACE since 2019. His main research interests include service robotics, distributed autonomous robotic systems, embodied brain science systems, and cognitive ergonomics.

Dr. Asama was the Recipient of the SICE System Integration Division System Integration Award for Academic Achievement in 2010, the JSME Award (Technical Achievement) in 2018, etc. He was the Vice President of RSJ from 2011 to 2012 and an AdCom member of IEEE Robotics and Automation Society from 2007 to 2009. He has been the President of IFAC since 2020. He has been a council member of the Science Council of Japan since 2017. He is also a Fellow of the JSME and RSJ.