Complementarity of Sensors and Weak Supervision for Defect Detection in Concrete Structures

Jun Younes Louhi Kasahara[1], Hiromitsu Fujii[2], Atsushi Yamashita[1] and Hajime Asama[1]

***Abstract*— The automation of concrete structure inspection methods such as the hammering test is highly desirable and critical, especially for social infrastructures such as tunnels and bridges. This is to ensure the safety of their users. Semi-supervised approaches have great compatibility with critical inspection methods since they allow to greatly reduce the workload on humans while still not removing them completely from the process, and thus providing some level of reassuring confidence. However, the performance of such semi-supervised approaches is conditioned by the correctness of the provided weak supervision by human and it can easily be imagined that, in practice, weak supervision will rarely be without errors. Therefore, the present paper proposes a method to complement weak supervision using sensor-provided information in order to both increase performance and mitigate the negative impacts of human errors. Experiments conducted in laboratory conditions using concrete test blocks in various configurations showed the effectiveness of the proposed method, returning better performance and higher robustness to errors in weak supervision.**

I. INTRODUCTION

Following recent catastrophic events such as the collapse of the Morandi bridge in Italy [1] or the collapse of the ceiling portion of the Sasago tunnel in Japan [2], increased attention has been put on the importance of concrete structure maintenance and inspection, especially when it comes to social infrastructures.

One popular method for such task is called the hammering test and consists of a human operator hitting the surface of the structure using a hammer and assessing from the returned impact sound the presence or absence of defects, as illustrated in Fig. 1. The popularity of this traditional method can be said to reside in its simplicity. However, one downside of the traditional hammering test is that it requires a human operator. Given the manpower shortage and the growing number of structures in need of testing, the automation of non-destructive testing methods such as the hammering test is highly desirable and has been the focus of several studies [3][4].

On the data analysis side, previous works about the automation of the hammering test employed machine learning, in both supervised and unsupervised approaches. [5] used time-frequency analysis and Ensemble Learning. [6] used ICA and Neural Networks. On the unsupervised side, clustering approaches using side information such as the position of hammering samples have been proposed in [7][8].

[1]Jun Younes Louhi Kasahara, Atsushi Yamashita and Hajime Asama are with the Department of Precision Engineering, The University of Tokyo, 113-8656 Tokyo, Japan. louhi@robot.t.u-tokyo.ac.jp

[2]Hiromitsu Fujii is with the Department of Advanced Robotics, Chiba Institute of Technology, 275-0016 Chiba, Japan.

Fig. 1. Hammering test being conducted by a professional inside a tunnel.

Supervised approaches usually exhibit superior performance compared to unsupervised learning approaches and this is extremely desirable, especially when it comes to critical tasks such as the inspection of social infrastructures where errors can potentially endanger the public. However, the performance of supervised methods is conditioned by the quality of the available training data. This is inherently unsuited for the task of defect detection since, while non-defect states can be known in advance and sampled, defect states by nature are unpredictable and can come in a potentially infinite number of configurations. In that sense, unsupervised approaches appear more practical but their lacking performance cannot be considered satisfactory.

There is another path, which is semi-supervised learning. In fact, semi-supervised or *weakly supervised* methods can be considered ideal for critical tasks such as inspection since they allows to partially automate processes while conserving human involvement, i.e., conserving human guarantees. Such approach was proposed in [9] with semi-supervised metric learning clustering, which used constraints, i.e., pairs of samples indicated to belong to the same class or not by a human. However, compared to the results obtained using clustering reinforced with position information as in [8], the detection performance was lacking. Moreover, the work proposed in [9] did not consider potential errors in the information provided by human.

Therefore, the aim of the present paper is to achieve better performance for semi-supervised approach to defect detection in concrete structures and to mitigate the performance loss incurred when the human providing weak supervision makes mistakes.

This is achieved by supporting human-provided weak supervision with sensor-based information and therefore, using the two datatypes to support and complement each another.

II. Overview of Proposed Method

Weak supervision provided by a human expert, i.e., answers to queries about similarity of pairs of hammering samples, provides crucial information about the desired feature space for defect detection. From the point of view of semi-supervised clustering, this would correspond to the constraint formulation as proposed in [10] and used to build appropriate feature space by methods such as Relevant Component Analysis (RCA) [11]. However, the benefit of building a feature space based on weak supervision greatly rests on the correctness of such provided information. It is not a scenario often considered for semi-supervised clustering methods but one can easily expect such human-provided information to comprise some errors. Therefore, some complementary information to support semi-supervision is required. Position information, i.e., the hit location of the hammer head, used in [8] is a good candidate. This is because, as a sensor-based information, it does not have the potential to incorporate errors such as the human-provided weak supervision, i.e., the position of a hamming sample might only be precise to a certain degree but can be reasonably assumed not to be inaccurate to the point of inversing the relative position of samples. An overview of the proposed method is shown in Fig. 2.

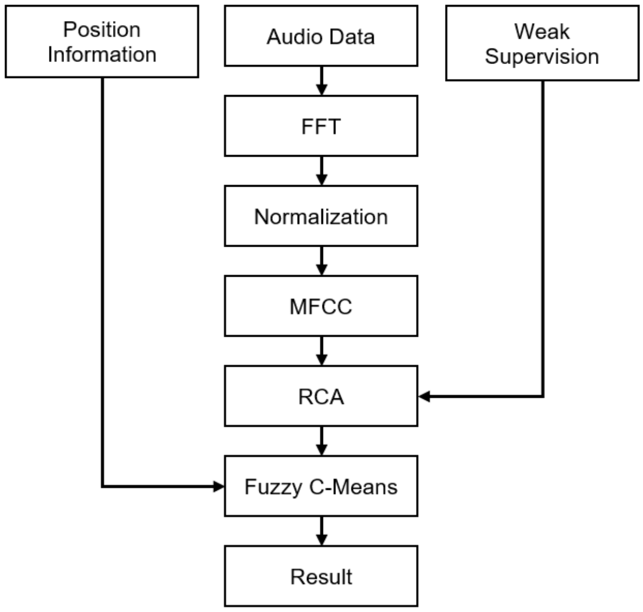


Fig. 2. Overview of the proposed method.

III. Audio Pre-Processing

Hammering samples are initially recordings of the impact sound of the hammer head on the surface of a concrete structure, i.e., time-series audio data. First, the Fourier spectrum of each hammering sample is obtained by FFT. Then, a zero-mean and unit variance normalization is conducted in order to remove potential discrepancies in hammering samples due to irregularities in the force used in the hammer strike. After this, MFCC feature vectors for each hammering sample are computed. MFCC is a feature vector most often encountered in speech recognition but its effectiveness for discrimination of hammering samples has been reported in [7]. In the remaining of this paper, a hammering sample's MFCC vector is simply noted $\mathbf{x}_i$.

IV. Semi-Supervised Feature Learning

RCA is a semi-supervised metric learning clustering method proposed in [11] and is basically a biased whitening transformation on the considered dataset. Given N hammering samples $\{\mathbf{x}_i\}$ and the associated set of must-link constraints, i.e., pairs of samples a human has indicated as similar, what are called chunklets are formed. Chuncklets $\{\mathcal{M}_l\}$ are groups of samples belonging to a same class by a human, i.e., some sort of *early clusters*, generated using the transitivity of must-links. With N_{chunklet} the number of such chunklets generated from weak supervision, and $\mathbf{m}_l$ the mean of elements in $\mathcal{M}_l$, the within-chunklet covariance matrix $\hat{\mathbf{C}}$ is computed as in (1). Then, the associated whitening transform is applied to the dataset, as in (2). RCA is therefore a biased whitening transform in the sense that the transformation is computed based on selected parts of the dataset and then applied to the whole dataset.

$$\hat{\mathbf{C}} = \frac{1}{N_{\text{total}}} \sum_{l=1}^{N_{\text{chunklet}}} \sum_{\mathbf{x}_i \in \mathcal{M}_l} (\mathbf{x}_i - \hat{\mathbf{m}}_l)(\mathbf{x}_i - \hat{\mathbf{m}}_l)^T, \qquad (1)$$

$$\mathbf{x}_i \rightarrow \hat{\mathbf{C}}^{-1/2}\mathbf{x}_i. \qquad (2)$$

Clustering to separate defect and non-defect hammering samples is then conducted on the feature space built using RCA, i.e., a feature space built on human-provided information.

V. Clustering with Position Information

For each hammering sample's MFCC vector $\mathbf{x}_i$ is associated position information $\mathbf{l}_i$, corresponding to the physical coordinates of the point on the structure that was hit with the hammer. Such sensor-based information can greatly help in the clustering in the audio space, even alone, as reported in [7].

The clustering is conducted using Fuzzy C-Means, an iterative, fuzzy clustering algorithm [12][13]. The flexibility allowed by fuzzy clustering, as opposed to the rigidity of crisp clustering methods such as K-Means, is important here to allow clustering in audio space while incorporating spatial information. After a random initialization step, Fuzzy C-Means alternates between centroid and fuzzy coefficient updates until movement of centroids have ceased. In the proposed method, the fuzzy coefficients update rule is composed of audio and spatial updates.

A. Audio Update Rule

The update of Fuzzy C-Means in the audio space corresponds to the traditionnal clustering update. For each hammering sample's MFCC vector $\mathbf{x}_i$ and for each K centroid $\mathbf{c}_j$, fuzzy membership coefficient u_{ij}, expressing the strengh of the sample's belonginess to the corresponding cluster, is computed. This is conducted as in (3), with m a parameter controlling the fuzziness of the system.

$$u_{ij} = \frac{1}{\sum_{r=1}^{N} \frac{\|\mathbf{x}_r - \mathbf{c}_i\|}{\|\mathbf{x}_j - \mathbf{c}_j\|}^{2/(m-1)}}. \tag{3}$$

B. Spatial Update Rule

Spatial information is used to enforce spatial compacity of clusters. This is because, given a tight enough hammering grid pattern, defects can be safetly assumed to be localized in the tested structure. Therefore, if two hammering samples are located in close proximity on the structure, they are more likely to belong to the same class. Such spatial correlation is injected into the clustering process in the audio space.

To this end, a spatial neighborhood $NB(\mathbf{l}_i)$ of each hammering sample is defined based on the position information as in (4). This is done as a disc of radius γ around the considered sample's position $\mathbf{l}_i$. Then, h_{ij}, a spatial estimation of the fuzzy coefficient based on this neighborhood, is computed as in (5), with $|NB(\mathbf{l}_i)|$ the cardinality of a neighborhood. This represents what the fuzzy coefficient of the sample should be according to other samples in its vicinity.

$$NB(\mathbf{l}_i) = (\mathbf{l}_i \in D \ \mid \ \|\mathbf{l}_i - \mathbf{l}_j\| \leq \gamma), \tag{4}$$

$$h_{ij} = \frac{1}{|NB(\mathbf{l}_i)|} \sum_{k \in NB(\mathbf{l}_i)} u_{kj}. \tag{5}$$

Finally, with p and q weighting exponents on each fuzzy components, a balancing step is conducted as in (6).

$$u_{ij} \rightarrow \frac{u_{ij}^p h_{ij}^q}{\sum_k u_{kj}^p h_{kj}^q}. \tag{6}$$

The centroid update rule remains unchanged from the regular Fuzzy C-Means. Conversion to a crisp clustering is done by maximum membership and cluster identification is conducted using the same process as in [8].

VI. EXPERIMENTS

Experiments were conducted in laboratory environment using concrete test blocks containing man-made defects. While such artificial defects may differ from naturally-occuring ones found in real inspection sites, they have the advantage that the exact specifications of the defect are precisely known, due to elaborate fabrication processes. Three cases were considered: Case 1, Case 2 and Case 3, shown in Fig. 4. Case 1 contains a single delamination-type defect. Case 2 is composed of two blocks, each containing one defect each. Case 3 is also composed of two blocks but one containing a delamination-type defect and the other a void-type defect. Those are the same blocks as the ones used in [8] and the number of clusters K is known as 2, 2 and 4, respectively.

The 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 sound board. MFCC were computed with 10 coefficients.

For weak supervision, 20 random must-links were provided by random selection of pairs and using the true labels. To generate errors in weak supervision, a portion of the must-links were replaced by random pairs of samples belonging to different classes. To obtain spatial information, the hammer head was painted in red and color-tracked with a camera. Neighborhood parameter γ was set empirically so that every hammering sample would have at least one neighbor. For the remaining parameters, $p = 1$, $q = 1$ for equal contribution of the two types of information and fuzziness parameter was left to the common setting $m = 2$. For all experiments, the Rand index [14], a commonly used performance measure based on groupings of sample pairs, was used for comparisons. The higher the value is, the better is the clustering output.

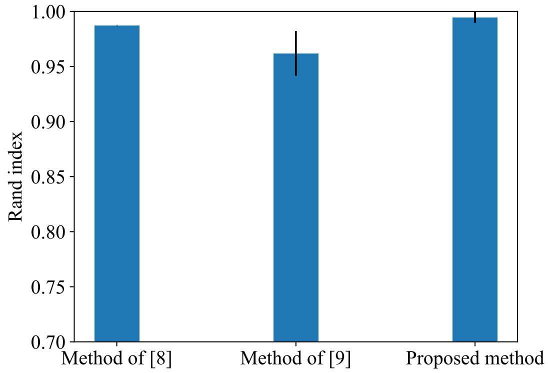
VII. RESULTS AND DISCUSSIONS

First, experiments without taking into consideration the possibility of human error were conducted, i.e., all provided must-links were correct.

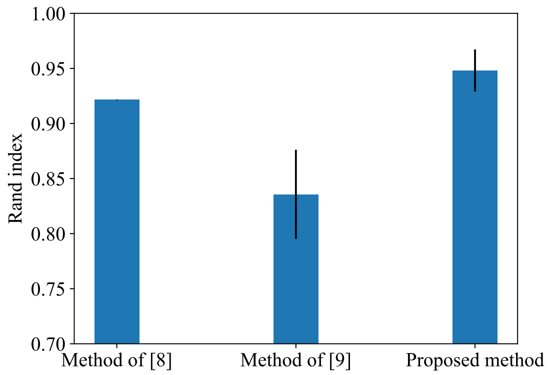
Corresponding results are reported in Fig. 3 for each considered case, which are pictured in Fig. 4. Across all cases, it can be seen that the method proposed in [9] using only weak supervision has the lowest performance. The method based on sensor-provided information proposed in [8] has higher average performance than the method proposed in [9]. Here, sensor-provided information is at a 1:1 ratio with the hammering data, i.e., there is position information for every hammering sample. On the other side, the method of [9] using weak supervision only has 20 pairs of samples declared as belonging to the same class as supporting information. In that sense, the fact that purely weak supervision having lower performance than the sensor-based approach is understandable.

The proposed method that employs both human-provided weak supervision and sensor-provided information has the best performance for all three considered cases. It is particularly remarkable for Case 1 where, out of the 20 runs with different sets of must-links and random seeding, a perfect result with a Rand index value of 1 was achieved six times. The fact that the proposed method employing both types of information outperforming the methods proposed in [8] and [9] strongly hints at a complementarity of the two datatypes and at a successful combination, where shortcomings of the weakly supervised component is covered by the sensor-based component and vis-versa.

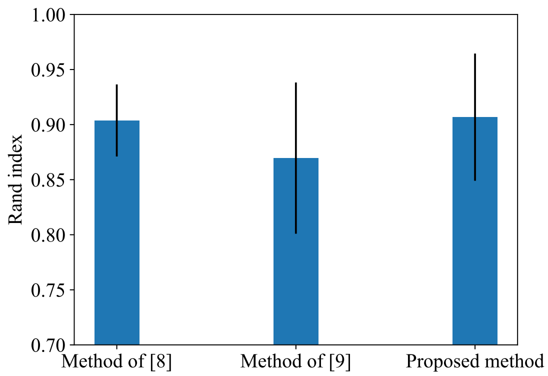
Actual clustering outputs of the methods proposed in [8], [9] and the proposed method of this paper are reported in



(a) Case 1: single delamination.



(b) Case 2: dual delaminations.



(c) Case 3: delamination and void.

Fig. 3. Performance in the case of no errors in weak supervision. Average performance over 20 iterations reported. 20 random must-links and random seeding for each iteration. Error bars corresponds to one standard deviation.

Figs. 5, 6, and 7.

For Case 1, the method using sensor-provided information returned a very good output, with a high Rand index value of 0.98. It can be seen in Fig. 5(a) that there is only a few misclassified samples, mainly on the bottom right-side of the defect area. This corresponds to the deepest part of the delamination, which is difficult to detect due to the impact sound being muffled. The weakly supervised method return a slightly worse output, with a Rand index of 0.94, shown in Fig. 6(a). The defect detection is not bad, except for a lump of misclassified samples on the left side of the block. Those are certainly due to border conditions being different at the edges of the blocks. It is worth noticing that the areas where each of these methods made errors are mainly different. This explains that the proposed method, using both weak supervision and sensor-provided information, was able to correctly classify samples in the two previously mentioned areas, as shown in Fig. 7(a). In fact, the result using the proposed method presented here has achieved a perfect Rand index of 1, which is remarkable.

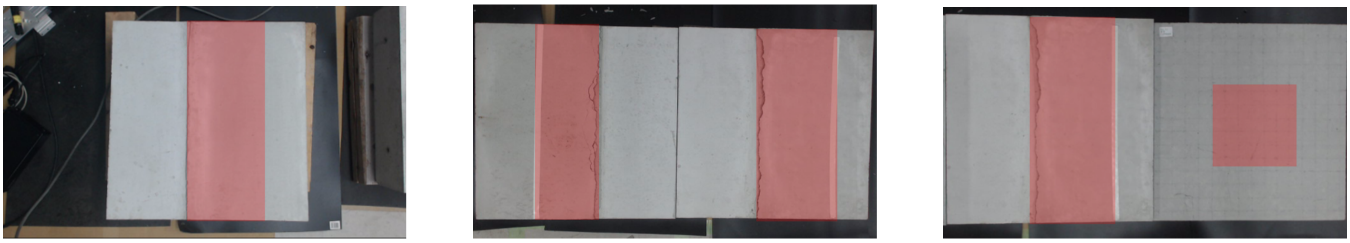
Case 2 is a more difficult scenario due to the presence of two distinct defects, albeit of the same type. The complementary of human-provided weak supervision and sensor-provided position information is less obvious here: as shown in Figs. 5(b) and 6(b), there are several samples that are misclassified in both outputs, such as those in the middle non-defect section and on the left side of the left defect. Since the method of [9] has a more significantly lower performance with a Rand index value of 0.84 against the method of [8], with a Rand index value of 0.92, their complementarity is less pronounced compared to Case 1. However, the proposed method successfully classifying samples misclassified in both previous methods, as shown in Fig. 7(b) with a Rand index value of 0.95, still hints at the fact that both types of information support each another.

For Case 3, the obvious point of difference is on the defect present on the middle of the right side block. While the method of [8] did manage to detect it, with some *overspill* due to border conditions at the edges of the block, the method of [9] completely missed that defect on this particular iteration, as shown in Figs. 5(c) and 6(c). The proposed method managed to output a better contour of that defect. Since position information is used after the weak supervision in the proposed method, this means that the use of position information succeeded at recovering the defect samples that would have been missed by a clustering algorithm purely on the audio space such as the one used in [9]. Furthermore, the final output is better than the output of the method proposed in [8]. This thanks to weak supervision having already conducted one portion of the separation defect/non-defect. Here too, the complementarity is strongly hinted.

Results about the impact of errors by human while providing weak supervision are presented in Fig. 8.

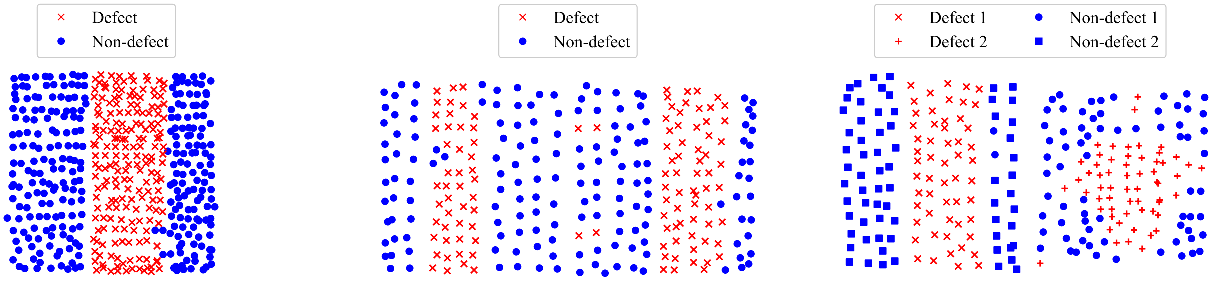
The first thing that can be noticed is that, as expected, the method proposed in [9] sees its performance decrease as the error rate in weak supervision increases. Since the method using position information proposed in [8] does not use weak supervision, it is not surprising that its performance is not correlated with the amount of errors in weak supervision. In fact, it is constant for Case 1 and 2. The fluctuations seen for Case 3 are due to the complexity of this case, mainly caused by its higher number of clusters and random seeding for Fuzzy C-Means.

It can be noted that the proposed method's ability to combine the best of the two types of information is beneficial for error rates in weak supervision under around 10% across all three considered cases. Beyond that threshold, the contribution of weak supervision is most likely too poor and starts to be detrimental to the contribution of position information. The performance loss caused by the presence of



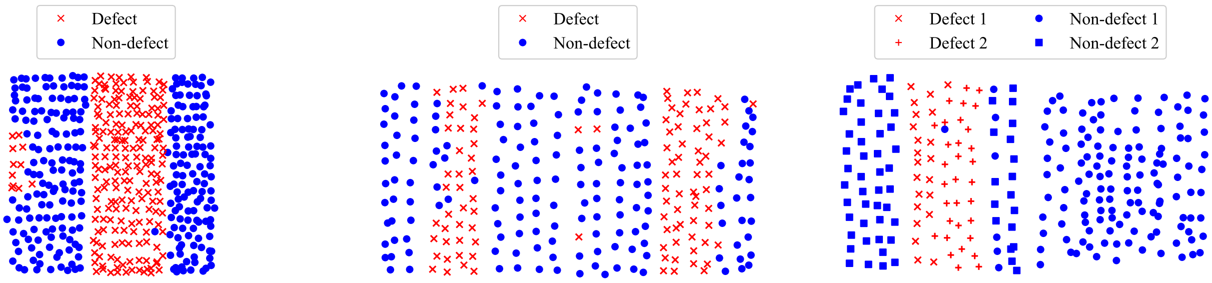
(a) Case 1: single delamination. (b) Case 2: dual delaminations. (c) Case 3: delamination and void.

Fig. 4. Concrete test blocks used in laboratory experiments. Light red area represent defect areas.



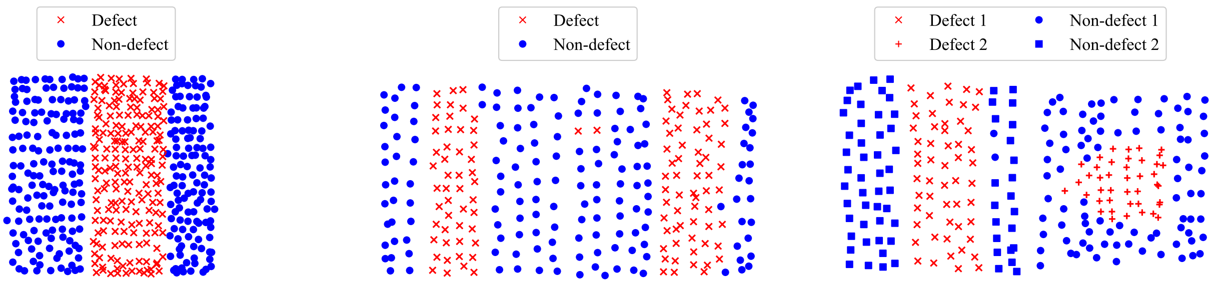
(a) Output of [8] on Case 1. (b) Output of [8] on Case 2. (c) Output of [8] on Case 3.

Fig. 5. Outputs of [8].



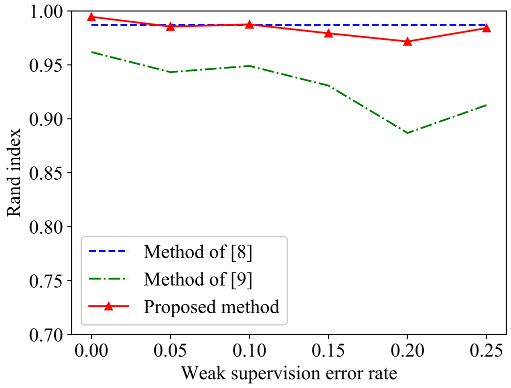
(a) Output of [9] on Case 1. (b) Output of [9] on Case 2. (c) Output of [9] on Case 3.

Fig. 6. Outputs of [9].

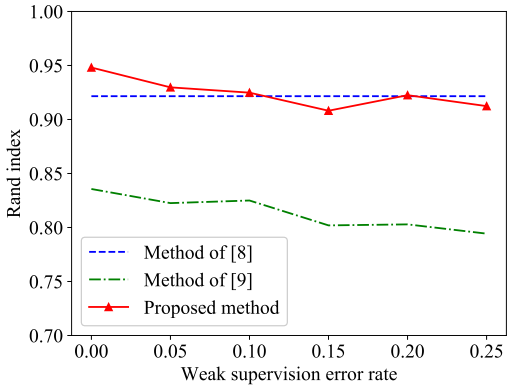


(a) Output of proposed method on Case 1. (b) Output of proposed method on Case 2. (c) Output of proposed method on Case 3.

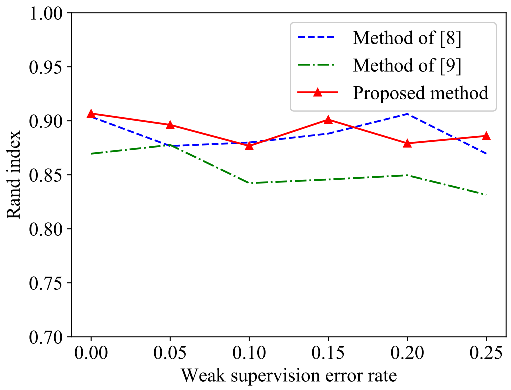
Fig. 7. Outputs of proposed method.



(a) Case 1: single delamination.



(b) Case 2: dual delaminations.



(c) Case 3: delamination and void.

Fig. 8. Performance with varying errors in weak supervision. Average over 20 iterations for each error rate reported. 20 random must-link and random seeding for each iteration.

false must-links is much more attenuated for the proposed method compared to the method of [9].

VIII. CONCLUSION

In this paper, a method using both human-provided weak supervision and spatial information, a sensor-provided information, was proposed. The proposed method successfully achieved higher performance than purely human-provided weakly supervised and purely sensor-based approaches for scenarios without errors in weak supervision. Regarding the impact of errors in weak supervision, the proposed method successfully suppressed them greatly, managing to ensure superior performance for up to about 10% errors.

As future work, we would like to conduct experiments in environment closer to actual concrete structure inspection sites, with natural defects. Furthermore, while the experiments conducted in the present paper showed the complementary of weak supervision and sensor-provided position information and effectively mitigated the potential performance loss due to errors in weak supervision, a process to filter out such errors using a sensor-provided information could potentially be able to withstand higher error rates in weak supervision.

ACKNOWLEDGMENT

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

REFERENCES

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

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

[3] Y. Takahashi, S. Nakamura, Y. Ogawa, and T. Satoh, "Velocity control mechanism of the under-actuated hammering robot for gravity compensation," in *Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC)*, 2017.

[4] S. Nakamura, Y. Takahashi, D. Inoue, and T. Ueno, "The variable guide frame vehicle for tunnel inspection," in *Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC)*, 2017.

[5] H. Fujii, A. Yamashita, and H. Asama, "Defect detection with estimation of material condition using ensemble learning for hammering test," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 3847–3854.

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

[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 *Proceedings of the International Symposium on System Integration (SII)*, 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," in *Robotics and Automation Letters*, vol. 3, no. 3. IEEE, 2018, pp. 2616–2623.

[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 *Proceedings of the International Conference on Consumer Electronics*, 2019 (accepted).

[10] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl, "Constrained k-means clustering with background knowledge," in *Proceedings of the International Conference on Machine Learning (ICML)*, vol. 1, 2001, pp. 577–584.

[11] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning a mahalanobis metric from equivalence constraints," in *Journal of Machine Learning Research*, vol. 6, no. Jun, 2005, pp. 937–965.

[12] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters." Taylor & Francis, 1973.

[13] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 1981.

[14] W. M. Rand, "Objective criteria for the evaluation of clustering methods," in *Journal of the American Statistical Association*, vol. 66, no. 336. Taylor & Francis Group, 1971, pp. 846–850.