

Defect Detection with Estimation of Material Condition Using Ensemble Learning for Hammering Test

Hiromitsu Fujii¹, Atsushi Yamashita¹ and Hajime Asama¹

Abstract—This paper introduces a new methodology of robotic hammering inspection for the maintenance of social infrastructures. In particular, the estimation of material defect conditions, such as delamination depth of concrete, is focused upon. Development of an automated diagnosis methodology is necessary for the maintenance of superannuated social infrastructures. The hammering test, which is an efficient inspection method, has attracted considerable attention in the context of automated inspection using robots. In this study, to apply the hammering test to robotic inspection, in which material conditions of infrastructures must be diagnosed in detail, an estimation method of the defect conditions is proposed, and an integration technique of plural classifiers for improving the inspection accuracy is introduced. Furthermore, an inspection system that can decrease the influence of the mechanical running-noise is implemented. Our experimental results using concrete test pieces demonstrate the effectiveness of the proposed method; the accuracy of the defect detection and defect condition estimation was validated.

I. INTRODUCTION

In recent years, the superannuation of social infrastructure has become a major problem involving concrete installations such as tunnels and bridges, particularly in Japan. Since the number of infrastructures that need inspection has increased considerably [1], it is extremely difficult to inspect all of them manually. Moreover, in many cases, the locations to be inspected, such as those in high and narrow places, are dangerous for human workers. Therefore, the development of an automated inspection system that can be installed on robots is strongly desired.

In the inspection work, not only the existence detection of material damage (*defect*) but also a detailed investigation of material defect conditions, such as crack depth, is necessary to find the extent to which the damage has progressed. For example, in general maintenance works of infrastructures, the inspection is divided into two stages, namely primary inspection and secondary inspection [2]. In a primary inspection, a rough inspection is conducted in a wide range. If some damage in progress is found, a secondary inspection is carried out in more detail.

Various types of non-destructive inspection methods have been developed [3]. In particular, the hammering test has attracted considerable attention in the context of automated inspection using robots. This method uses dedicated sticks called percussion sticks or inspection hammers in several forms (Fig. 1). The workers distinguish material damage

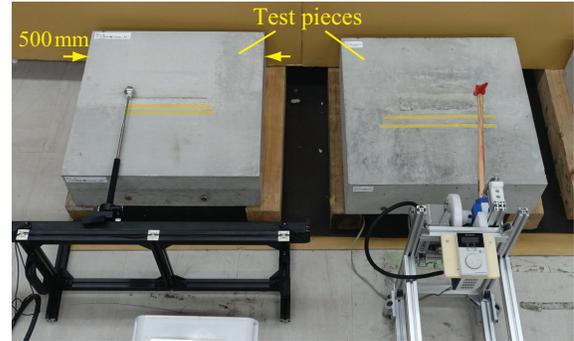


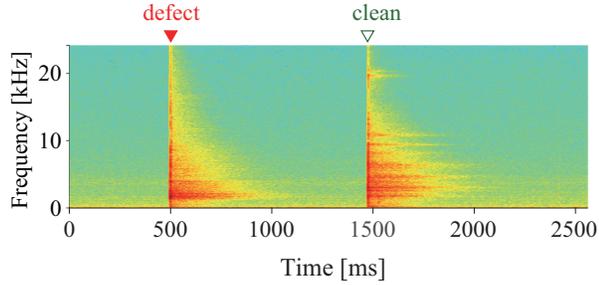
Fig. 1. Prototypes of hammering modules and concrete test pieces. The machine on the left side equipped with a rolling hammer and a motor slider, is the robot for the first screening whose main purpose is to find the defect effectively. The other on the right side is the robot equipped with a swing hammer for the secondary inspection in which the material condition is diagnosed in detail. Experiments in this study were conducted using the latter robot.

by identifying the difference in hammering sounds without any contact sensor, such as accelerometer. This method has been used for a long time and is still widely used for manual inspection because of the accuracy of the resulting diagnosis and ease of execution. Owing to these advantages, the hammering test is considered suitable for robotic implementation and is desired to be applied to both stages of inspection. In particular, the defect condition estimation by using a hammering test with robots has a high demand for an effective execution of the secondary test.

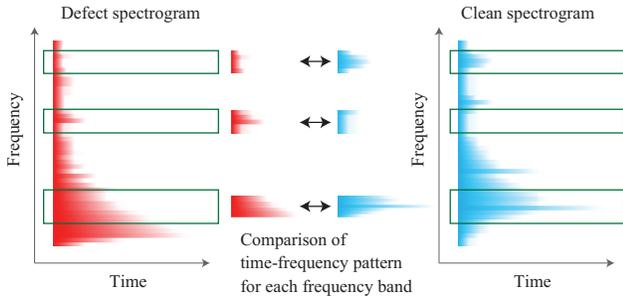
Thus far, numerous studies on the hammering test have been reported [4]–[12]. The analysis-based studies [6], [9], [11] have extracted effective features to detect material defects; however, they did not mention the robotic automation of their techniques. With respect to robotic implementation [4], [7], [8], they took measures against the influence of problems of the actual environment, such as traffic vibrations. However, most of them focused only on defect detection and did not adequately discuss the classification of defects.

Our research group has also developed an automated inspection method to detect material defects by using the hammering test [5], [10], [12]. In a previous work, it has been already confirmed that the proposed boosting based algorithm based on a frequency pattern analysis can be used for constructing a noise-robust defect detector against an environmental acoustic change [10], [12]. In this study, focusing on the estimation of the defective material condition, the accurate defect detector that performs the function of a clustering-based multi-label estimation is proposed. In particular, it detects defects by evaluating the similarity of the

¹H. Fujii, A. Yamashita, and H. Asama are with the Department of Precision Engineering, Faculty of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. {fujii, yamashita, asama}@robot.t.u-tokyo.ac.jp



(a) Comparison of spectrogram between a *defect* signal and a *clean* signal. The spectrogram patterns change according to the conditions of the material. For example, in this case, the clean signal has a peak that lasts at a high frequency of around 20 kHz.



(b) Pattern matching by using sub-band similarity in time-frequency variations, which can be treated as a two-dimensional pattern. The proposed method integrates plural classifiers that deal with the time series variations of their own specific frequency bands.

Fig. 2. Schematic of the template matching of time-frequency pattern.

sub-band in the time-frequency domain. In order to improve the accuracy of defect detection and defective condition estimation, an ensemble learning-based integration method of plural classifiers is proposed. Furthermore, an inspection system with mechanical noise avoidance of self-running is constructed. The experiments demonstrate that the proposed system can detect delamination detection and can estimate the delamination depth with high accuracy.

II. DEFECT DETECTOR BASED ON TIME-FREQUENCY ANALYSIS

In this section, the proposed method for detecting material defects is presented. By applying an ensemble learning technique, the proposed method integrates plural detectors (weak learners) that can deal with their own respective frequency subband. In the following parts, let symbol D denote variables regarding *defect*, and let symbol C be used for describing variables regarding *clean*, which means defect-free.

A. Template Matching of Time-Frequency Pattern

In the hammering test, inspection workers diagnose the material condition by relying on various types of information, such as pitch and change in the hammering sound. The short-time Fourier transform (STFT) is applied to represent the features of the hammering sound. STFT is a method for time-frequency transformation, which has been used for many types of nondestructive inspections [13]. The hammering

sound can also be analyzed via a time-frequency dimensional representation by considering specific frequency levels and their time series changes. These features of the hammering sound change according to the conditions of the material. An example of a spectrogram comparison between a *defect* signal and a *clean* one is shown in Fig. 2(a). In particular, in this case, the clean signal has a peak that lasts at a high frequency of around 20 kHz.

In this study, weak learners discern defects of the material by analyzing the shape of the time-frequency distribution, which can be treated as a two-dimensional pattern. In particular, the similarity in time-frequency variations is evaluated in the frequency sub-band as shown in Fig. 2(b). In order to classify signals robustly without the influence of the sound pressure, a zero-mean normalized cross-correlation (ZNCC) that is extended by introducing the weights of the frequency components is applied. The evaluation function $S(\mathbf{A}, \mathbf{u}, \mathbf{x})$ is represented as follows:

$$S(\mathbf{A}, \mathbf{u}, \mathbf{x}) = \frac{\sum_{j=1}^J \sum_{k \in \mathcal{K}} u(k) \tilde{A}_{(j,k)} \tilde{x}_{(j,k)}}{\sqrt{\sum_{j=1}^J \sum_{k \in \mathcal{K}} u(k) \tilde{A}_{(j,k)}^2} \sqrt{\sum_{j=1}^J \sum_{k \in \mathcal{K}} u(k) \tilde{x}_{(j,k)}^2}}, \quad (1)$$

$$\tilde{A}_{(j,k)} = A_{(j,k)} - \bar{A}, \quad (2)$$

$$\tilde{x}_{(j,k)} = x_{(j,k)} - \bar{x}, \quad (3)$$

where \mathbf{A} represents the template vector calculated with the frequency data of the training set. The template vector \mathbf{A} is calculated considering the weights $w^{(i)}$ of training samples $\mathbf{x}^{(i)}$ ($i = 1, \dots, I$), where I denotes the count of the training samples. They are defined as $A_{(j,k|k \in \mathcal{K})} = \sum_{i=1}^I w^{(i)} x^{(i)}_{(j,k)}$. The weights $w^{(i)}$ can be obtained in the learning step of boosting for integrating weak learners. Let \bar{A} and \bar{x} denote the mean values of \mathbf{A} and \mathbf{x} , respectively. The variable j denotes the index of the FFT windows that continues in the time direction, and the constant parameter J denotes the count of the FFT windows. The set \mathcal{K} represents the index set of the frequency components used in the classification of \mathbf{x} , which corresponds to the frequency sub-band that the weak learner deals with. The index k denotes each frequency index in \mathcal{K} . The vector \mathbf{u} consists of the weights of the frequency components included in \mathcal{K} , which represents the feature vector of a hammering sound. The function evaluates the similarity between template vector \mathbf{A} and input vector \mathbf{x} . The range of $S(\mathbf{A}, \mathbf{u}, \mathbf{x})$ is $[-1, 1]$ and is the same as that of the general ZNCC. The parameters \mathcal{K} and \mathbf{u} must be designed for the defect detector, and the details of procedures to decide these parameters \mathcal{K} and \mathbf{u} are described in our previous work [12].

In Particular, for a momentary hammering sound, the difficulty arising in signal matching is that the shapes of the frequency distribution change according to the sampling timing. The resulting signal of STFT includes the frequency signals of any sampling timing. Therefore, from the aspect of learning, this approach has an advantage of training dataset preparation, and it leads to an improvement of the diagnostic

accuracy. Furthermore, it leads to the capability of online inspection. The diagnosis can be conducted in any timing by performing a single FFT that can be processed fast, since it learned from the dataset containing the spectrum of various time phases.

B. Discriminant Function of Clean-Defect

In the weak learner, defect detection has a function of classifying *clean* and *defect*. With respect to the performance of each learner, since the proposed method aims to construct the defect detector having both high accuracy and high robustness by integrating plural learners, the whole diversity is prioritized to the respective accuracy. This approach can also avoid overfitting the training dataset.

Let ${}^cS(\mathbf{x}) := S({}^cA, {}^c\mathbf{u}, \mathbf{x})$, and let ${}^D S(\mathbf{x}) := S({}^D A, {}^D \mathbf{u}, \mathbf{x})$, which are obtained by a similarity evaluation for *clean* template cA and *defect* one ${}^D A$ by calculation according to eq. (1). In the case that the frequency sub-band \mathcal{K} is selected properly, similarity value pair $[{}^cS(\mathbf{x}) \ {}^D S(\mathbf{x})]^T$ on the training sample \mathbf{x} is distributed in the cS - ${}^D S$ space, as shown in Fig. 3(a). That is, if sample \mathbf{x} belongs to clean class C , ${}^cS(\mathbf{x})$ should be large, being close to 1.0. Similarly, if sample \mathbf{x} belongs to defect class \mathcal{D} , ${}^D S(\mathbf{x})$ should be large. Consequently, the clean samples and defect ones can be discriminated in the cS - ${}^D S$ space.

Let us introduce a linear discriminant function to classify the clean and defect samples as follows:

$$\mathbf{m}^T \mathbf{S} + \theta = 0, \quad (4)$$

where $\mathbf{S} = [{}^cS \ {}^D S]^T$, \mathbf{m} denotes the coefficient vector, and θ denotes the bias, both of which must be designed for each weak learner. Detection of defects by \mathbf{x} that each weak learner $h(\mathbf{x}) \in \{-1, 1\}$ performs is as follows:

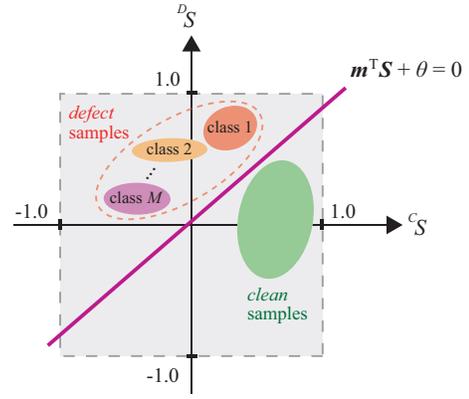
$$h(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{m}^T \mathbf{S} + \theta \geq 0 \\ -1 & \text{otherwise} \end{cases}. \quad (5)$$

For example, when $h(\mathbf{x}) = 1$, the sample is defective.

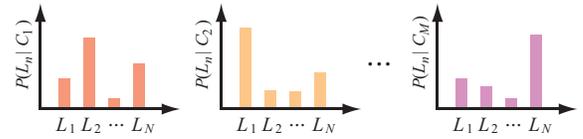
Inner parameters \mathbf{m} and θ are optimized in order to appropriately classify the training samples of clean and defect in the cS - ${}^D S$ space. The criterion of classification is to maximize the margin between C and \mathcal{D} . This problem is well known as the training of a linear support vector machine. In this study, considering the case where both classes cannot be completely discriminated with eq. (4), this optimization problem is formulated with a soft margin and is solved by using the coordinate descent method.

C. Condition Estimation Using Mean Shift Clustering

The samples in the training dataset are labeled with *clean* and *defect*. The preliminary label of new samples can be discriminated by the method discussed until Section II-B. Additionally, defect samples are labeled with more detailed information about their condition for the secondary inspection. It is adequately assumed that the samples having the same label are located close to one another in the cS - ${}^D S$ space. In order to estimate the labels of the new samples, the distribution of training samples is clustered in advance



(a) Schematic of classification by linear discriminant function. The linear discriminant function is calculated to maximize the margin between *clean* samples and *defect* ones. Furthermore, *defect* samples are clustered by using the mean shift algorithm. Each cluster includes the sound samples labeled according to the additional detailed label, such as delamination depth. Based on the clusters, the label of a new sample is estimated as shown in Fig. 3(b).



(b) Estimation of material condition in detail. The condition of the new data is estimated from the posterior probability distributions $P(L_n | C_m)$ of additional multi labels $L_n (n = 1, \dots, N)$ that can be obtained from the training samples in the clusters $C_m (i = 1, \dots, M)$.

Fig. 3. Defect detection and detailed defective condition estimation for secondary inspection, which are introduced in a single classifier. In the proposed method, the whole detector integrates the plural classifiers equipped with these functions, and each classifier performs in its own frequency band (Section II-D).

(Fig. 3(a)). How reliably the members of a cluster have the label is expressed as a posteriori probability distribution $P(L_n | C_m)$ by calculating the counting statistics of the training dataset as shown in Fig. 3(b), where $L_n (n = 1, \dots, N)$ denotes labels indicating detailed material condition attached in advance and $C_m (m = 1, \dots, M)$ denotes clusters composed from training the samples. In this study, a mean shift-based clustering algorithm [14] is applied to compose the cluster in the cS - ${}^D S$ space.

In the diagnosis, first, the label of a new sample is estimated. The new sample \mathbf{x}_{new} is examined to find the cluster that it belongs to. The cluster of sample C_{new} is chosen to satisfy the following:

$$C_{\text{new}} = \arg \min_{C_m} \|\bar{\mathbf{p}}_{\{C_m\}} - \mathbf{p}_{\mathbf{x}_{\text{new}}}\| \quad (m = 1, \dots, M), \quad (6)$$

where $\bar{\mathbf{p}}_{\{C_m\}}$ denotes the mode of cluster C_m and $\mathbf{p}_{\mathbf{x}_{\text{new}}} = [{}^cS(\mathbf{x}_{\text{new}}) \ {}^D S(\mathbf{x}_{\text{new}})]^T$. Accordingly, the weak learner outputs a posteriori probability distribution of the condition labels as $P(L_n | \mathbf{x}_{\text{new}}) := P(L_n | C_{\text{new}})$. Next, the label of the new data L_{new} can be estimated according to $\arg \max_{L_n} P(L_n | \mathbf{x}_{\text{new}})$.

D. Sub-band Detector Integration Using Ensemble Learning

A weak learner diagnoses the material condition by focusing on its own specific frequency sub-band. In order

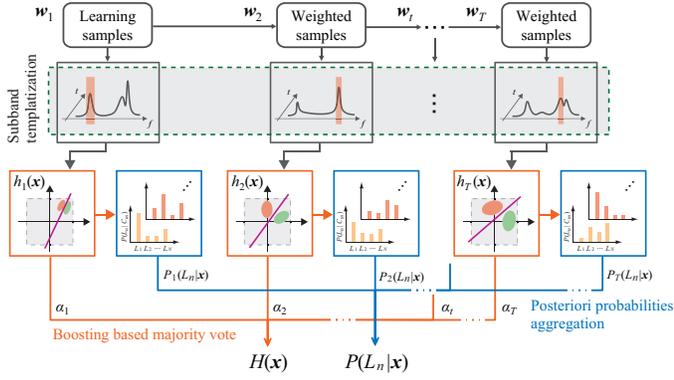


Fig. 4. Schematic of proposed defect detector. Each weak learner has frequency templates in its own frequency sub-band. In the diagnosis, defect detection and detailed condition estimation are conducted simultaneously. Whole detector aggregates their answers, outputting it as the final result.

to deal with various material defects, which have different characteristics in the frequency domain, improvement of both accuracy and robustness is quite significant. For this purpose, we propose an integration of weak learners by using ensemble learning techniques.

A schematic representation of the proposed detector is shown in Fig. 4. With respect to material defect detection, for accurate diagnosis, it is necessary for weak learners to obtain their variety by focusing on the sub-band different from one another. In order to construct such weak learners, the proposed method generates weak learners in sequence according to a boosting algorithm [15] and integrates them into an whole detector (strong learner). This approach was validated in our previous work [10], [12]. The strong learner $H(\mathbf{x})$ can be expressed as follows:

$$H(\mathbf{x}) = \sum_{t=1}^T \alpha_t \text{sign}[h_t(\mathbf{x})] / \sum_{t=1}^T \alpha_t, \quad (7)$$

where α_t denotes the confidence coefficient of each weak learner as computed by the error ratio. The output of the strong learner is expressed as a type of defective score by normalizing it in the range of $[-1, 1]$. That is, the higher (the closer to 1.0) $H(\mathbf{x})$ is, the more is the hammered diagnostic target suspected of a defect. For example, in the case that the plus or minus sign of $H(\mathbf{x})$ is adopted as a defect criterion, the defect detection can be treated as an output of the binary classification $H^*(\mathbf{x}) \in \{-1, 1\}$ as follows:

$$H^*(\mathbf{x}) = \begin{cases} 1 & \text{if } \text{sign}[H(\mathbf{x})] \geq 0 \\ -1 & \text{otherwise} \end{cases}. \quad (8)$$

In this paper, the estimation methodology of the detailed material condition is proposed. In our boosting-based framework of the whole detector, each weak learner $h_t(\mathbf{x})$ outputs the posteriori probability distribution of condition labels $P_t(L_n|\mathbf{x})$ ($n = 1, \dots, N$), as described in Section II-C. The probability distributions of all learners are summed up. The label $L^*(\mathbf{x})$ of the new sample \mathbf{x} is estimated as the one that has the highest probability as

$$L^*(\mathbf{x}) = \arg \max_{L_n} \sum_{t=1}^T P_t(L_n|\mathbf{x}). \quad (9)$$

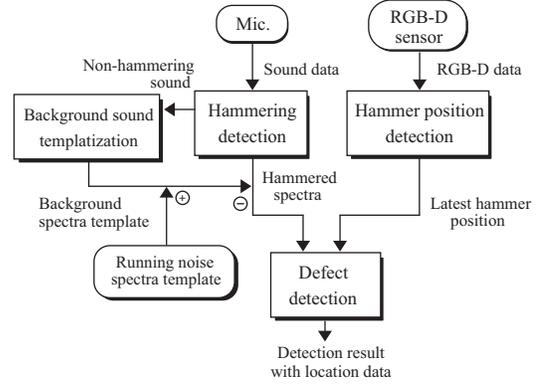


Fig. 5. Schematic of our defect detection system. The system has functions of noise influence avoidance with hammering sound detection and struck point detection by using an RGB-D sensor.

This aggregation of the probabilities of a weak learner is similar to bagging techniques [16], such as the Random Forest. The proposed method is characterized by the fact that what is selected by bootstrapping is not training samples but frequency components. It leads to a diversity of probabilistic outputs not by the difference in the decision trees but by the various frequency sub-bands that each weak learner is assigned to.

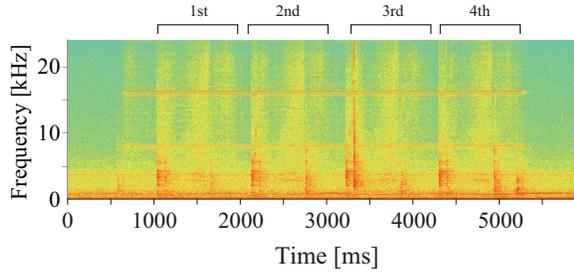
III. AUTOMATED INSPECTION SYSTEM

The proposed method is implemented as an automated hammering system using an automated module. A signal pipeline of the detection system is shown in Fig. 5. In this section, processing noise avoidance and three-dimensional measurement of a struck point are briefly introduced.

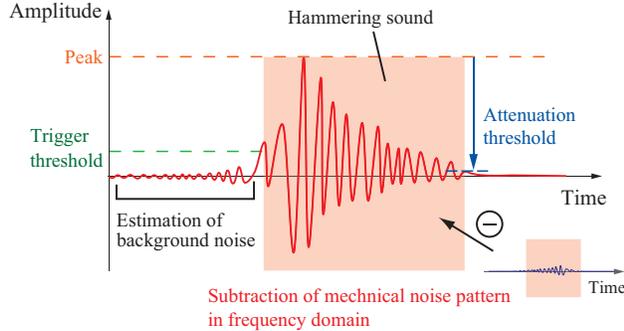
A. Self-Noise Robust Detection in Robotic Hammering

The automated module for secondary inspection (shown on the right side of Fig. 1) is equipped with a swing hammer. The robot swings down the inspection hammer by a magnetic mechanism. In our previous work [12], the effect of environmental noises was successfully reduced by subtracting the background spectrum by using a conventional and simple but efficient method [17]. When using mechanical modules, it must also be considered that running noises from the module itself influence the result of the inspection. An example of a spectrogram of the running noise from the module during a striking motion without contact against an inspection target is shown in Fig. 6(a). In order to avoid the influence of the mechanical noise, the time-frequency pattern is templatized in advance and is subtracted in the frequency domain during the hammering.

As shown in Fig. 6(b), the rise time of the hammering sound is detected by a trigger. After the triggered point and during a certain time window, signals with a larger amplitude than the attenuation threshold are considered reverberations. The attenuation threshold value is set in advance by considering an amplitude damping factor towards the maximum signal peak of the duration.



(a) The spectrogram of self-running robotic noise of a module. Mechanical noise of the module itself exists during hammering. For example, it can be confirmed from sharp changes in the spectrogram color at around 8 kHz and 17 kHz.



(b) Hammered sound detection. Rise signal and its duration are detected as a hammered sound by a trigger. Spectrum of the background sound can be estimated during the non-hammered period.

Fig. 6. Avoidance of various noise influence of robotic hammering.

B. Hammered Position Detection

To automate inspection works, not only existence detection of a defect but also specification of the defect location is relevant. For example, in Japan, the inspection report must include defect locations and the quantitative scale of the defect area by law. In order to measure the locations and sizes of defects, the struck position of hammering is estimated by using an RGB-D sensor, by which the RGB texture with the depth value in every pixel can be obtained (Fig. 7). First, the largest plane is detected as the target plane of inspection by applying RANSAC to the depth information obtained from the sensor, since the hammering target of concrete structures is planar. Second, when the hammer head gets closest to the target plane, the struck point is estimated as the foot of a perpendicular line from the point detected as the hammer head. The hammer head is the part of the hammer that is painted in red in Fig. 7. In the experiment of this study, inspection hammer is actually painted with red and its head is detected by image processing using the RGB texture information obtained from the sensor. The three-dimensional position of the hammer head can be estimated by using the coordinate of the head in the captured image, the depth of the same pixel, and implicit parameters of the sensor.

IV. EXPERIMENTS

With respect to various types of concrete defects, delamination is a particularly serious one since pieces that delaminate from a wall frequently fall down and damage

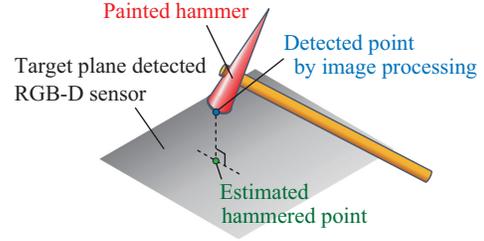


Fig. 7. Struck position detection by using RGB-D sensor. The head of the inspection hammer is painted in order to detect it by image processing. The struck position is estimated by the three-dimensional position of the tip and the target plane, both of which can be obtained from the RGB-D sensor data.

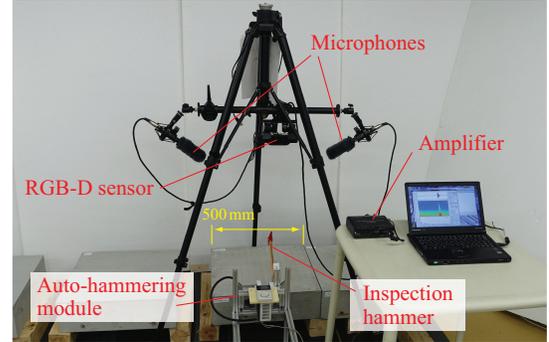


Fig. 8. Experimental devices. The proposed system consists of the auto-hammering module, microphones, an RGB-D sensor, and a signal processor that implements the proposed algorithm.

cars and injure pedestrians. Delamination is a complex phenomenon that can arise, for example, from closing cracks or expansions of rusted reinforcing iron. Therefore, while the automation of accurate delamination detection is highly desirable, it is a challenging task. In addition, particularly in secondary inspection works, it is also needed to measure the depth of delamination, which means how deeply a crack reaches from a superficial layer. By measuring the location of delamination with depth information, the status of defect progression and its distribution can be estimated. In the experiments, we have therefore particularly focused on the detection of delamination and the estimation of its depth.

A. Experimental Settings

The experimental equipment used in our study is shown in Fig. 8. Two condenser microphones and an amplifier were used for capturing the hammering sounds. The resolution and the sampling rate were 24 bit and 48.0 kHz, respectively. As a diagnostic tool, we used an inspection hammer, which is generally used for the inspection of concrete infrastructures. The hammer head was painted in red for detection. The hammer head was detected by the image processing procedure (Section III-B) with an RGB-D sensor, ASUS Xtion Live Pro. The diameter and the weight of the hammer head were 12.4 mm and 0.1 kg, respectively. The hammer was installed on an auto module to equalize the force of striking and the time interval of the strike motion.

Concrete test pieces with an artificial delamination area such as those shown in Fig. 1 were crafted for the purpose

Table I
THE SPECIFICATIONS OF CONCRETE TEST PIECES. ALL THESE HAVE THE SAME VOLUME
($W \times D \times H = 500\text{mm} \times 500\text{mm} \times 150\text{mm}$).

Test pieces	Crack entry angle	Extension size	Max depth of crack
for learning (TP _{L1})		clean (no crack)	
for learning (TP _{L2})	15 deg	200 mm × 71 mm	19 mm
for learning (TP _{L3})	30 deg	200 mm × 64 mm	37 mm
for learning (TP _{L4})	45 deg	200 mm × 53 mm	53 mm
for learning (TP _{L5})	15 deg	200 mm × 142 mm	38 mm
for validation (TP _{V1})	30 deg	200 mm × 128 mm	74 mm

of both learning and validation. Specifications of the test pieces used in this experiment are given in Table I. One clean (defect-free, TP_{L1}) test piece and four test pieces with a delamination defect (TP_{L2}, TP_{L3}, TP_{L4}, and TP_{L5}) were used for learning. In addition, one defect test piece (TP_{V1}) was used for validation. The depth of delamination was altered by changing the entry angle of the crack in each test piece. For example, the specification of TP_{V1} is shown in Fig. 9.

The training dataset for learning was collected by using the test pieces (TP_{L1}–TP_{L5}). The dataset includes *clean* and *defect* samples. As an experiment of the detailed condition estimation, the defect samples are labeled additionally with two labels as *shallow* and *deep* with respect to their own delamination depth. The training dataset consists of 2,775 sound samples including 1,257 delamination samples and 1,518 clean samples. The additional labels were also set manually in advance; the *shallow* label denotes the defect whose depth is up to 30 mm and the *deep* label denotes one over 30 mm, since the depth of 30 mm is one of the significant criteria for finding how seriously the delamination has progressed. An example of the correspondence between the additional label and the physical status is shown in Fig. 9 (red masked area indicates *shallow* area and blue area, *deep*). The additional labeled samples include 663 *shallow* samples and 855 *deep* ones.

With respect to the parameters of the defect detector, the window size for FFT was 1,024 corresponding to around 21 ms. The count of weak learners was 128.

B. Experimental Results

1) *Results of Ensemble Learning*: Figure 10(a) shows a result example of defect detection by a weak learner. The horizontal axis represents the similarity with a clean template and the vertical axis represents one with defect samples. *Clean* samples and *defect* samples are classified by a linear discriminant function in the cS - $^D S$ space, although they are not separated completely because of the adoption of soft-margin maximization. Here, note that the depth labeled defect samples are located near each other in the class with the same label. In Fig. 10(b), a clustering result of the defect samples is shown. Each cluster has different labeled samples; thus, the weak learner obtained the posteriori probability distribution $P(L_{\text{depth}}|C_m)$ ($m = 1, \dots, M|M = 4$ in this case).

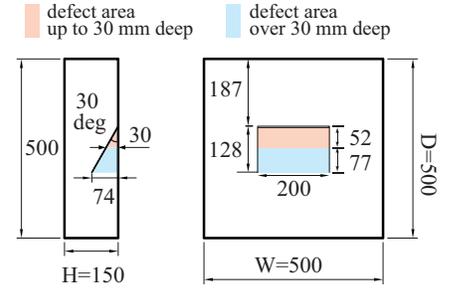


Fig. 9. Schematic of test piece TP_{V1} for validation. A slant crack intrudes into the surface. The farther from the crack entrance it is, the deeper the crack reaches.

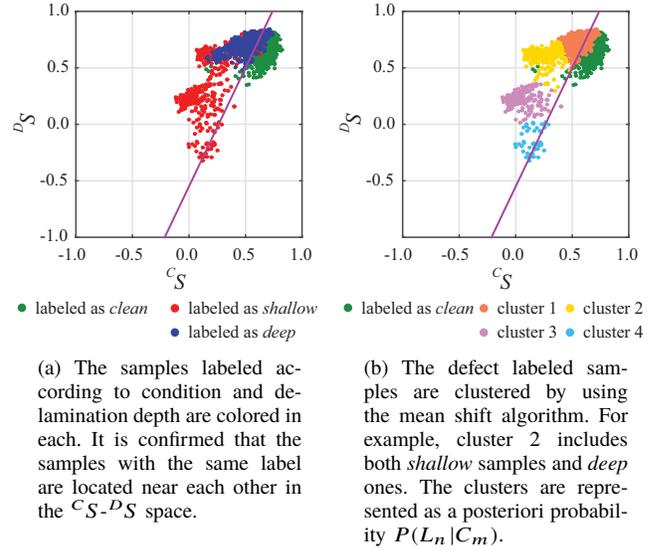
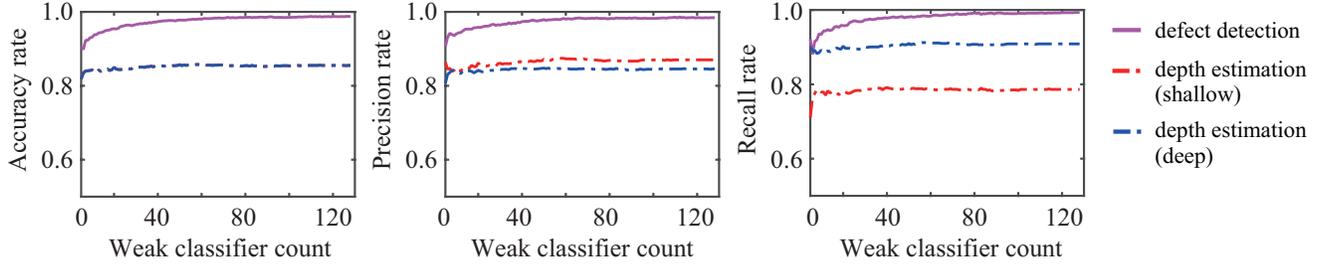


Fig. 10. An example of classification of training samples by a weak learner. *Clean* and *defect* samples are classified by a linear discriminant function.

The result of the ensemble learning of the proposed defect detector is shown in Fig. 11. In order to validate the effect of the integration of plural learners, the relation between the performance of detection and the count of weak learners integrated into the whole detector is shown. As performance indices, accuracy (Fig. 11(a)), precision (Fig. 11(b)), and recall (Fig. 11(c)) values are evaluated as the average of the results by 5-fold cross validation for the training dataset. In all the figures, the horizontal axis represents the count of the weak learners and the vertical axis represents the performance rate. The purple solid line indicates the result of delamination defect detection and the red and blue dotted lines indicate the estimation results of *shallow* and *deep* depth labels, respectively. By integrating plural learners, the performance of defect detection can be increased effectively. With respect to the depth estimation, although the performances increase gradually by combining weak learners up to around 40, subsequently, the performances get saturated. This is considered to be caused by the ambiguity of the labels set manually. The samples with depth around the border of 30 mm were difficult to classify into one class because of their topological and acoustic similarity. However, it has

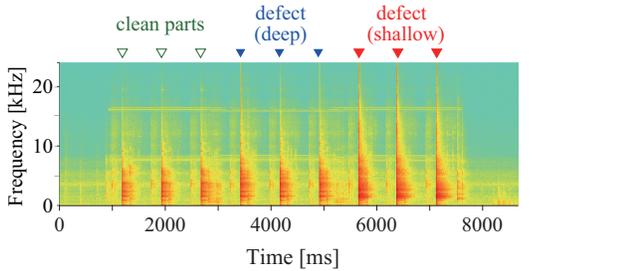


(a) The result of *accuracy*. The index shows to what extent the answers of the detector are correct.

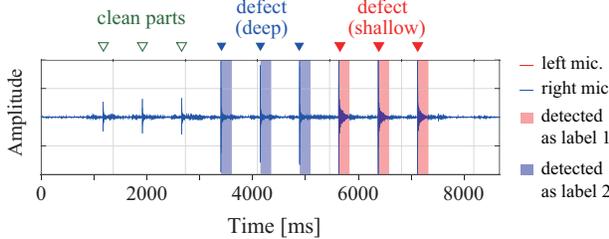
(b) The result of *precision*. The index shows to what extent the samples the detector classifies as defect are truly the defect.

(c) The result of *recall*. The index shows to what extent the actual defect samples can be correctly detected by the detector.

Fig. 11. The results of detector performance evaluation, in which the results of defect detection and depth estimation are shown. The horizontal axis represents the number of weak learners that constitute the whole detector. The vertical axis represents the index rate that indicates the performance of the detector. These results were evaluated as the average of the 5-fold cross validation of the training dataset. The charts indicate that the proposed detector can accurately detect delamination defect; furthermore, it can correctly estimate the depth that the crack attains. In particular, the accuracy of defect detection (Fig. 11(a)) is increased by integrating plural weak learners. In the same chart, the accuracy results of the delamination depth estimation (red and blue dotted lines) coincide because true positive and true negative can be alternated between the two defect labels (*shallow* and *deep*).



(a) The spectrogram of hammering sound by being struck by auto module in order of three trials against a clean part, three trials against a defect part labeled *deep*, and three trials against a defect part labeled *shallow*. The noise mentioned in Fig. 6(a) influences the sound analysis of hammering.



(b) The result of defect detection and classification. The detected hammering sound is correctly diagnosed. The clean part was not detected as a defect, and the defect parts were detected with the correct label according to their own crack depth.

Fig. 12. The result of defect detection and delamination depth estimation.

been confirmed that the proposed method can construct, by integrating plural weak learners, a highly reliable defect detector that can detect the delamination defects with high accuracy and can estimate their depth labels with over 80% accuracy.

2) *Results of Automated Defect Detection*: After the construction of the defect detector, an experiment of automated diagnosis was conducted against the test piece TP_{V1} for validation, whose specification is shown in Fig. 9. The test piece was not used for learning but only for the validation.

The automated module struck the surface of the test piece by order, and the diagnostic results were shown on the display.

An example of hammering signals is shown in Fig. 12(a), and the result of the diagnosis is shown in Fig. 12(b). Figure 12(a) shows the spectrogram measured during hammering; the horizontal axis represents time [ms], the vertical axis represents frequency [kHz], and the depth of color indicates the amplitude strength in the frequency domain. Figure 12(b) shows the diagnostic result of the mic-input signal in the time domain; the horizontal axis represents time [ms], the vertical axis represents the amplitude of min-input. In the experiment, the automated module struck the surface nine times continuously in 8 s. The nine trials consisted of three hammer strikes against clean parts, three hammerings against delamination parts with the *deep* label, and three against delamination parts with the *shallow* label, in order. The areas emphasized by the half-tone background with the red color indicate the time intervals corresponding to the detection of the defects with the *shallow* label, and those with the blue color indicate the defects with the *deep* label. The defects were detected, and their depth were correctly estimated. The statistical result is shown in Fig. 13(a) as a receiver operating characteristic (ROC) curve. The 256 samples were evaluated, and the parameter of the detector regarding the balance between the false positive and the true positive was the output value of the strong learner $H(x)$, which is mentioned in eq. (7). With respect to the area under the curve (AUC), which was 0.98 close to 1.0, the performance of the diagnosis was highly accurate.

In the same experiment, defect position estimation was conducted. A scene of the experiment is shown in Fig. 13(b), and the result is shown in Fig. 13(c). In both the figures, the same topological coordinate system of XY is shown correspondingly. In Fig. 13(b), the detected position of the inspection hammer head is colored in deep red. The defect areas labeled *shallow* and *deep* were emphasized by the half-tone background with red and blue, respectively. In

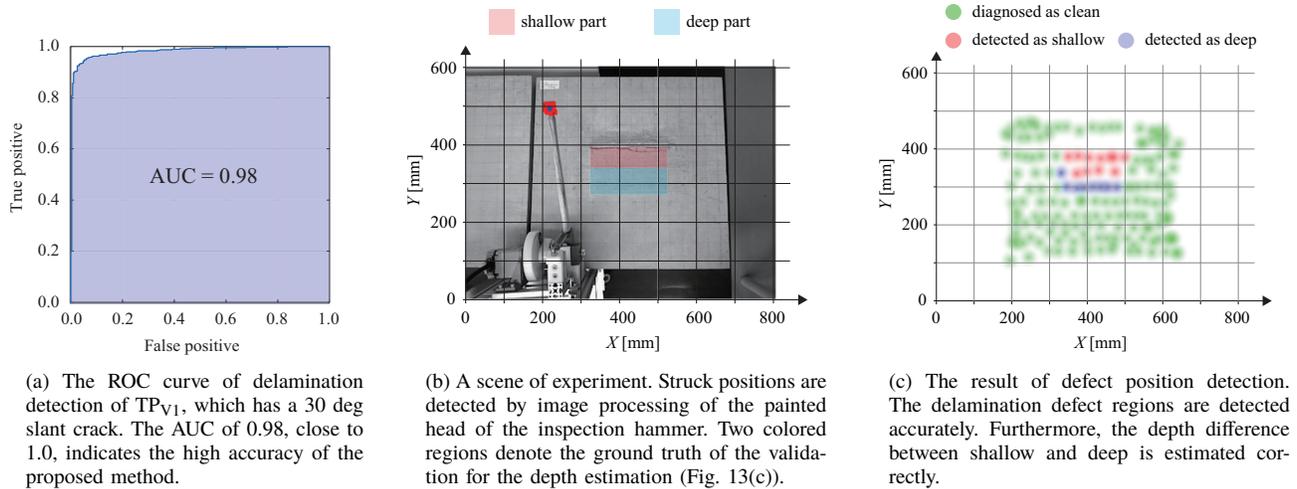


Fig. 13. The result of defect position detection and delamination depth estimation by using the auto module equipped with the proposed method.

Fig. 13(c), hammered points are marked by colored circles, the colors (green, red, and blue) indicate the diagnostic results as *clean*, *shallow delamination defect* and *deep delamination defect*, respectively. Thus, it is confirmed that the proposed inspection system can accurately detect the position of defects with their depth.

V. CONCLUSION

In this study, in order to construct the defect detector for the secondary inspection of a hammering test, we proposed and demonstrated an ensemble learning-based defect detector that can detect defects with high accuracy and can estimate the detailed material condition with additional labels. The experimental results that showed the proposed method can accurately distinguish signals and can estimate the material conditions by using additional labels.

As a future work, we plan to validate the performance of the proposed method against a variety of additional labels, and to apply this system to an actual environment.

ACKNOWLEDGMENTS

This work was in part supported by Council for Science, Technology and Innovation, “Cross-ministerial Strategic Innovation Promotion Program (SIP), Infrastructure Maintenance, Renovation, and Management” (funding agency: NEDO), Grant-in-Aid for JSPS Fellows 26-9039, and Institute of Technology, Tokyu Construction Co., Ltd.

REFERENCES

- [1] Takada, R., Oishi, N., “Priority Issues of Infrastructure Inspection and Maintenance Robot: A Part of COCN 2012 Project ‘Disaster Response Robot and its Operational System’,” Humanitarian Technology Conference (R10-HTC), 2013 IEEE Region 10, pp. 166–171, 2013.
- [2] Asakura, T. and Kojima, Y., “Tunnel Maintenance in Japan,” Tunneling and Underground Space Technology, Vol. 18, No. 2, pp. 161–169, 2003.
- [3] Hoła, J., Bień, J., Sadowski, Ł. and Schabowicz, K., “Non-destructive and Semi-destructive Diagnostics of Concrete Structures in Assessment of Their Durability,” Bulletin of the Polish Academy of Sciences Technical Sciences, Vol. 63, No. 1, pp. 87–96, 2015.
- [4] Suda, T., Tabata, A., Kawakami, J. and Suzuki, T., “Development of an Impact Sound Diagnosis System for Tunnel Concrete Lining,” Tunneling and Underground Space Technology, Vol. 19, Issue 4–5, pp. 328–329, 2004.
- [5] Yamashita, A., Hara, T. and Kaneko, T., “Inspection of Visible and Invisible Features of Objects with Image and Sound Signal Processing,” Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3837–3842, 2006.
- [6] Sonoda, Y. and Fukui, Y., “A Basic Study on Hammering Tests of Deteriorated Concrete Structures,” 35th Our World in Concrete and Structures, 100035063, 2010.
- [7] Inoue, F., Doi, S., Ishizaki, T., Ikeda, Y. and Ohta, Y., “Study on Automated Inspection Robot and Quantitative Detection of Outer Tile Wall Exfoliation by Wavelet Analysis,” Proceedings of International Conference on Control, Automation and Systems 2010, pp. 994–999, 2010.
- [8] Zhang, G., Harichandran, R. S. and Pradeep, R., “An Automatic Impact-Based Delamination Detection System for Concrete Bridge Decks,” NDT & E International, Vol. 45, No. 1, pp. 120–127, 2012.
- [9] Ye, J., Iwata, M., Kobayashi, T., Murakawa, M., Higuchi, T., Kubota, Y., Yui, T. and Mori, K., “Statistical Impact-Echo Analysis Based on Grassmann Manifold Learning: Its Preliminary Results for Concrete Condition Assessment,” Proceedings of 7th European Workshop on Structural Health Monitoring, pp. 1349–1356, 2014.
- [10] Fujii, H., Yamashita, A., and Asama, H., “Automated Diagnosis of Material Condition in Hammering Test Using a Boosting Algorithm,” Proceedings of the 2014 IEEE Workshop on Advanced Robotics and its Social Impacts, pp. 101–107, 2014.
- [11] Igual, J., Salazar, A., Safont, G. and Vergara, L., “Semi-Supervised Bayesian Classification of Materials with Impact-Echo Signals,” Sensors, Vol. 15, Issue 5, pp. 11528–11550, 2015.
- [12] Fujii, H., Yamashita, A. and Asama, H., “Improvement of Environmental Adaptivity of Defect Detector for Hammering Test Using Boosting Algorithm,” Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 6507–6514, 2015.
- [13] Cocconcetti, M., Zimroz, R., Rubini, R. and Bartelmus, W., “STFT Based Approach for Ball Bearing Fault Detection in A Varying Speed Motor,” Condition Monitoring of Machinery in Non-Stationary Operations, Springer Berlin Heidelberg, pp. 41–50, 2012.
- [14] Cheng, Y., “Mean shift, mode seeking, and clustering,” Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 8, pp. 790–799, 1995.
- [15] Freund, Y. and Schapire, R. E., “A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting,” Journal of Computer and System Sciences, Vol. 55, Issue 1, pp. 119–139, 1997.
- [16] Breiman, L., “Bagging predictors,” Machine learning Vol. 24, No. 2, pp. 123–140, 1996.
- [17] Boll, S. F., “Suppression of Acoustic Noise in Speech Using Spectral Subtraction,” IEEE Transaction on Acoustics, Speech and Signal Processing, Vol. 27, No. 2, pp. 113–120, 1979.