# Inspection of Visible and Invisible Features of Objects with Image and Sound Signal Processing

Atsushi Yamashita, Takahiro Hara and Toru Kaneko
Department of Mechanical Engineering
Shizuoka University
3–5–1 Johoku, Hamamatsu-shi, Shizuoka 432–8561, Japan
Email: yamashita@ieee.org

*Abstract*— In this paper, we propose a new method that can inspect visual and non-visual features of objects simultaneously by using image and sound signal processing techniques.

A method for discriminating a property of an object with the use of generated sound when striking it with a hammer is called a hammering test. This method can investigate non-visual features of objects such as inner structure of objects, *e.g.*, the existence of defects and cracks inside objects. However, this method depends on human experience and skills. In addition, if we perform this test over a wide area of objects, it is required to manually record hammering positions one by one.

To solve these problems, this paper proposes a hammering test system consisting of two video cameras that can acquire image and sound signals of a hammering scene. The shape of the object (visual feature) is measured by the image signal processing from the result of 3-D measurement of each hammering position, and the thickness or material (non-visual feature) is estimated by the sound signal processing in time and frequency domains.

The validity of proposed method is shown through experiments.

## I. INTRODUCTION

In this paper, we propose a new inspection method of visual and non-visual features of objects by using image and sound signal processing techniques.

By striking an object and hearing generated sound, humans can recognize the difference in thickness, the quality of materials, or the existence of defects. An inspecting method using the generated sound when we strike an object with a hammer is called a hammering test. It has an advantage that it performs non-destructive and simple investigation.

On the other hand, it has disadvantages as follows:

1) This test lacks objectivity because it depends on individual experience and skills.
2) This test inspects only the existence of defects inside an object. The external shape of the object is checked by human and is not measured precisely.
3) It is very time consuming and difficult to create database of results because it is required to manually record hammering positions one by one if we perform this test over a wide area of an object.

In order to solve the problem 1), it is necessary to analyze the hammering sound numerically by the computer. As examples of such research, there are studies on inspection of concrete cracks [1]–[5]. These studies only evaluate the difference between defects and normal parts, and don't execute the
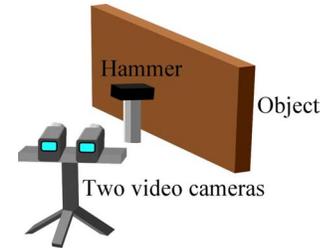


Fig. 1.   Overview of hammering test.

hammering test to estimate non-visual features quantitatively. A method using ultrasonic waves is also employed to detect the depth of cracks [6]–[8]. However, the transmission-and-reception probe must be contacted the object around cracks whose positions are estimated before the depth profiling. Therefore, the wide-range inspection is also difficult by using ultrasonic waves. The damage detection method by using the image processing technique is also proposed [9]. However, this method cannot treat with non-visual defects.

As to the problems 2) and 3), the shape of objects can be measured quantitatively by cameras instead of human's eyes. The database of the results can be easily created by the computer with image processing techniques.

Therefore, we propose a new inspection method of visual and non-visual features of objects by using image and sound signal processing techniques. A hammering test system consisting of two video cameras that can acquire image and sound signals of a hammering scene is constructed (Fig. 1). The shape of the object (visual feature) is measured by the image signal processing from the result of 3-D measurement of each hammering position by triangulation, and the thickness or material (non-visual feature) is estimated by the sound signal processing in time and frequency domains.

The composition of this paper is detailed below. In Section II, the procedure of our proposed method is mentioned. In Section III explains discrimination maps, and Section IV discuss the method of the hammering test. Section V shows the experimental results and Section VI describes conclusions and future works.

## II. PROCEDURES

The procedure of our method is divided into two processes.

- **Pre-processing**
    1) Sound signal processing
    2) Clustering of non-visual features
    3) Generation of discrimination map
- **Hammering test**
    1) Estimation of non-visual features by sound signal processing
    2) Estimation of visual features by image signal processing
    3) Display of results by merging information about visual and non-visual features

In pre-processing, we aim at generating a discrimination map that represents the difference of non-visual features of the object by using colors. This map is generated from the features of sample points before the hammering test by striking the typical points of the object with a hammer. In this study, the human operator must take notice that the hammer is perpendicular to the object surface when striking the object.

In hammering test, non-visual features are estimated by comparing extracted features from sound signals while striking the object with the features of the discrimination map. Visual features are measured from image signals of two cameras by stereo measurement. The results are displayed by merging information about visual and non-visual features.

## III. PRE-PROCESSING

In pre-processing, the discrimination map is generated. Since the captured signals are a complex of sound and image signals, we should divide the signals into two signal sets. Separation is realized by reading the signals structure unit by unit sequentially, and contiguous sound signals and contiguous image signals are composed, respectively.

### A. Sound Signal Processing

In the sound signal processing, the non-visual features such as the thickness and the quality of the material of objects is acquired. The sound signals are examined both in time and frequency domains in this study.

*1) Processing in Time Domain:* In time domain analysis, the decay time of a sound wave is extracted. It is necessary to find the time $T_s$ when the sound begins. The time when the sound pressure level becomes suddenly large defines $T_s$. The smallest value of $t$ (time) that satisfies the following equation is regarded as $T_s$.

$$|y(t + \Delta t_1)| - |y(t)| > L_1, \tag{1}$$

$$\left| \frac{y(t + \Delta t_1)}{y(t)} \right| > L_2, \tag{2}$$

where $y(t)$ is the amplitude of sound wave at time $t$, $\Delta t_1$ is the time interval unit, and $L_1$, $L_2$ are threshold values, respectively.

In indoor acoustics, the decay time is defined as when the amplitude decreases by 60dB, *i.e.* it decreases to $1/1000$.
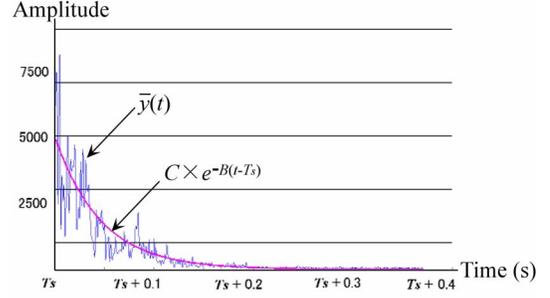


Fig. 2. Example of raw sound wave and its approximation curve.

However, analyzing the hammering sounds in such cases as in outdoor environments or under the conditions where ambient noises exist, it is difficult to apply this criterion. Therefore, we define the decay index $B$.

We assume that averaged amplitude $\overline{y}(t)$ can be represented by the following equation, where $\overline{y}(t)$ is the average of the absolute values of $y(t)$ within a short duration $\Delta t_2$.

$$\overline{y}(t) = C \times e^{-B(t-T_s)}, \tag{3}$$

where $t > T_s$.

Then, we obtain the indexes $B$ and $C$ in the above equation by using the least mean square method. The index $B$ given by this step is used for the sound signal discrimination in later process. The index $C$ is not used because it strongly depends on the hammering power. An example of the result is shown in Fig. 2.

*2) Processing in Frequency Domain:* In the frequency domain analysis, the Fast Fourier Transform (FFT) is executed to obtain spectral data. At first, average frequency $\overline{f}$ is calculated as follows:

$$\overline{f} = \sum_{i=1}^{N} \frac{f_i A_i^2}{A_s^2}, \tag{4}$$

$$A_s^2 = \sum_{i=1}^{N} A_i^2, \tag{5}$$

where $A_i$ is the $i$-th amplitude, $f_i$ is the $i$-th frequency, and $N$ is the number of data, respectively.

The average frequency $\overline{f}$ is easy to become the same value when several sound signals themselves are different from each other. Therefore, we use the spectral distribution of the sound signal. A frequency feature value $V_f$ (frequency variation) is calculated by (6), where a spread in frequency dependence is considered. An example of the FFT result is shown in Fig. 3.

$$V_f = \frac{1}{A_s} \sqrt{\sum_{i=1}^{N} A_i^2 (f_i - \overline{f})^2}. \tag{6}$$

### B. Clustering

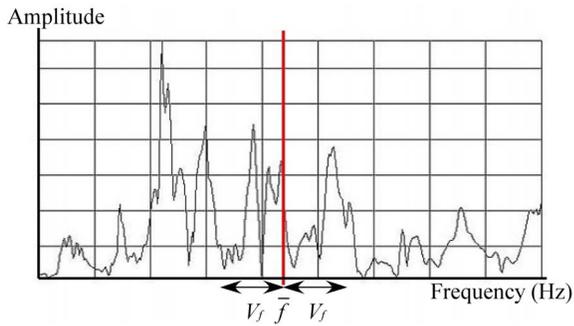Clustering process divides non-visual features into multiple classes (groups).
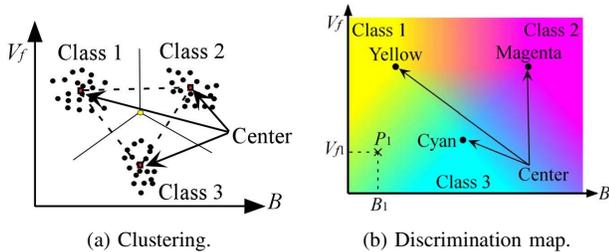
Fig. 3.    Example of frequency spectrum with FFT.



(a) Clustering.                (b) Discrimination map.

Fig. 4.    Clustering and discrimination map in $B$-$V_f$ diagram.

The decay index $B$ and the frequency variation $V_f$ are calculated by the sound signal processing for each sample signal, and we obtain the $B$-$V_f$ diagram having $B$ and $V_f$ on the horizontal and vertical axis, respectively. Clustering process is executed on the $B$-$V_f$ diagram by using ISODATA (Iterative Self Organizing Data Analysis Technique A) method [10]. The ISODATA clustering method can decide the number of classes and the center position of each class in the feature space automatically.

This technique is based on randomly choosing initial cluster centers, or means. These initial cluster centers are updated in such a way that after a number of cycles they represent the clusters in the data as much as possible. The ISODATA algorithm circumvents the problem by removing "redundant" clusters. Whenever a cluster centre is not assigned enough samples, it may be removed. In this way one is left with a more or less optimal number of clusters. We adopt the variety of ISODATA method proposed in [11].

After finishing the clustering process, the sound signals in each hammering points can be divided into several classes in the $B$-$V_f$ diagram and the center position of each class is obtained (Fig. 4(a)).

### C. Generation of Discrimination Map

A discrimination map is created in advance, in order to distinguish non-visual features of objects such as the thickness and the quality of the material.

The value that expresses the difference in non-visual features is given for each center position. In this study, the color values are given for the center positions. The color value at each position in $B$-$V_f$ diagram is calculated according to the
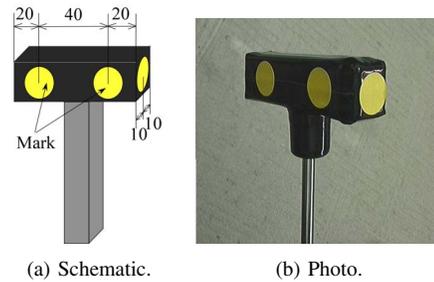


(a) Schematic.                (b) Photo.

Fig. 5.    Overview of the hammer.

distance from center positions. Therefore, the color values mean the probability of belonging to a certain class. In other words, the neutral color means the rate of likelihood of each class. We define this color map as a discrimination map. Figure 4(b) shows an example of the discrimination map when the number of the class is three, and the color of each center is yellow, magenta, and cyan, respectively.

## IV. Hammering Test

The visual and non-visual features of the object are estimated by striking the surface of the object one by one.

### A. Estimation of Non-Visual Features

Once the discrimination map is generated, we can estimate the non-visual features of any portion of the object by comparing its values $B$ and $V_f$ to the map. For example, $B_1$ and $V_{f1}$ are the feature values from the sound signal when striking the object in Fig. 4(b), the color at the point $P_1$ (green: mixed color of yellow and cyan) represents the non-visual feature of that hammering point.

Here, it should be noted that each hammering gives a non-visual feature estimate and also its location on the object by the procedure of image signal processing.

### B. Estimation of Visual Features

It is necessary to measure the 3-D positions of hammering points to reconstruct the 3-D shape of the object. Therefore, 3-D information of the hammering points is acquired from the image signals of two cameras by triangulation [12].

Since the image is taken from behind the hammer, the exact hammering position is hidden by the hammer itself. Therefore, in order to know the position accurately, it is required to calculate by integrating information about figures in the image, camera parameters including viewing distance and angles, and a 3-dimensional shape model of the hammer.

In this study, marks are attached to the hammer to estimate its position and orientation easily (Fig. 5). The hammering points on the surface of the object can be estimated by measuring the marks' positions of the hammer.

At first, the image frame corresponding to the event at time $T_s$ is extracted from the image sequence file. The synchronized signal between two cameras is not needed because $T_s$ can be extracted from each camera's image and this time is perfectly same with each other.
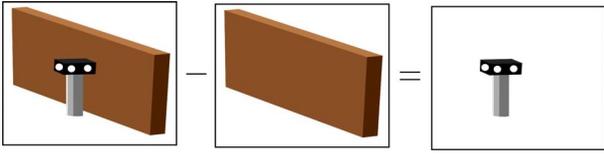
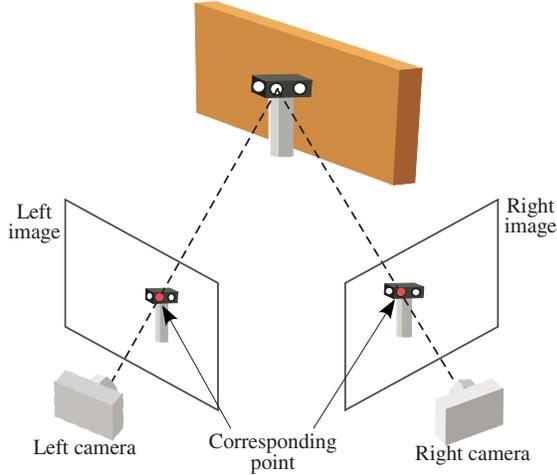Fig. 6. Extraction of the hammer from the image.


Fig. 7. Stereo measurement.

The extraction method of marks on the hammer is explained as follows. In the first step, only the hammer is extracted from the image. This can be realized by comparing the image at the time $T_s$ and the image in which there are no hammer (Fig. 6). The regions where the difference between above two images is larger than a certain threshold value are extracted as the candidate regions of the hammer. There are large contrast between the marks and the rest of the hammer. Therefore, we can extract the pixels belonging to the marks. After extracting them, the opening operation (erosion and dilation) is executed for eliminating small noises. After that, each mark can be distinguished from each other by the labeling operation. Finally, the center of each labeling region is calculated as the centroid of these pixels.

Above process is carried out for two images from two cameras, and corresponding points between left and right images can be gained. The 3-D coordinate value of each mark on the hammer is measured from these corresponding points with triangulation (Fig. 7). In other words, the 3-D position of each mark is obtained with the stereo measurement by using two cameras. The 3-D position and the normal vector of each hammering point on the object can be estimated by reconstructing the shape of the hammer.

### C. Display of Results

In order to examine non-visual or visual features over the whole object, the process mentioned above needs to be done very densely on the object. However, this is difficult in practice and, if possible, takes a lot of time.

Therefore, we hammer at discrete points and interpolate points between each hammering point with interpolation.

As to the 3-D measurement results, the reconstruction of the object's shape is executed. When several points are on the same plane, we construct the plane from these points by the regression analysis. When they are not on the same plane, the object's shape is expressed by triangle patches.

As to the non-visual features, the linear interpolation decides the object's surface colors that indicate the difference of the non-visual features.

## V. EXPERIMENTS

In experiments, the objects were hammered with almost the same strength. Two digital video cameras were used so that sound and image signals could be easily captured.

External and internal parameters of each camera were calibrated by using planar pattern on which surface circular patterns are drawn [13].

The signals downloaded to a computer were saved as the AVI file, and this file was divided into the sound signal file and the video signal file. Parameters of the sound and image files were set as follows.

- Sound signal
  - Sampling frequency: 44.1kHz
  - Number of channels: 1 (monaural recording)
  - Number of bits: 16 bits
- Image signal
  - Resolution: 720×480pixel
  - Frame rate: 29.97frame/s

We set parameters as follows: $L_1 = 20000$, $L_2 = 5$, $\Delta t_1 = \Delta t_2 = 5$msec.

On executing FFT, the Hanning window was used, and its size was 16384 (=2^14), which was given by the recording period 0.372sec and the sampling frequency 44.1kHz. The frequency feature value $V_f$ was calculated within the frequency band from 100Hz to 1kHz, which gave specific frequency distribution.

### A. Thickness

A wood board shown in Fig. 8 was examined about the thickness of the object. The board size is 600mm by 600mm. The board is divided into three areas that have three thickness values, 312mm (Class 1), 47mm (Class 2), and 12mm (Class 3). An example of an image of a hammering scene is shown in Fig. 9(a), and the enlarged image of extracted hammer and marks is shown in Fig. 9(b).

Examples of the waveform and the frequency spectrum are shown in Fig. 10. Figure 10(a), (b), (c) show that there is a differences between the waveforms and the frequency distributions for the thickness of the object.

The thickness in each point on the board is inspected by using the discrimination map shown in Fig. 11(a).

By hammering several times at 7 by 8 points on the board, each thickness was estimated. Figure 11(b) shows the estimates for the actually hammered positions and their colors.
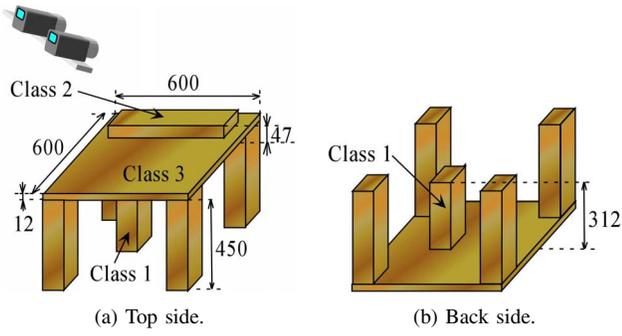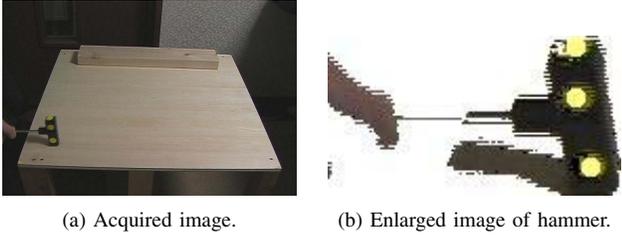
(a) Top side.
(b) Back side.

Fig. 8. Object I.


(a) Acquired image.
(b) Enlarged image of hammer.

Fig. 9. Extraction of hammer's marks.


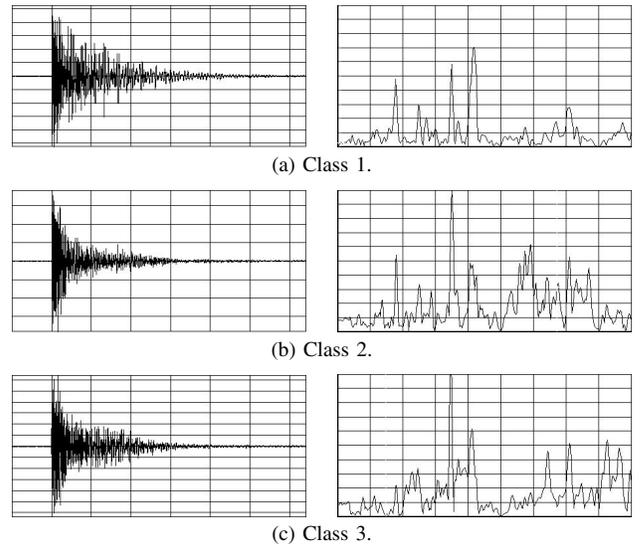(a) Class 1.


(b) Class 2.


(c) Class 3.

Fig. 10. Examples of raw sound wave (left figure, vertical axis: amplitude, horizontal axis: time) and frequency spectrum with FFT (right figure, vertical axis: amplitude, horizontal axis: frequency).
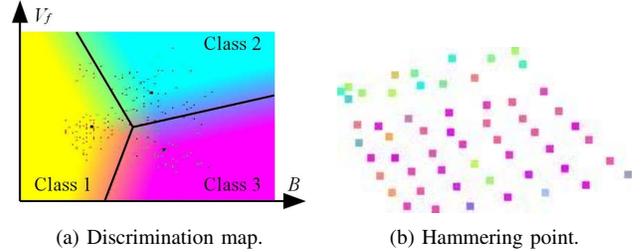

(a) Discrimination map.
(b) Hammering point.

Fig. 11. Discrimination map and 3-D positions and colors of each hammering point for object I.

Figure 12(a) shows the result by interpolation. By comparing of this result with the actual shape and thickness distribution shown in Fig. 12(b), it can be said that the experimental result coincided with the actual one.

### B. Quality of Material

Figure 13 shows the object that consists of four materials; wood (Class 1), cork (Class 2), plastic (Class 3), and styrene foam (Class 4). These four kinds of quality of the materials were pasted up. In the experiment, a while cover is overlaid, since the difference in the quality of the material cannot be distinguished from a color from images.

The quality of the material in each point on the board is inspected by using the discrimination map shown in Fig. 14(a). Figure 14(b) shows the estimates for the actually hammered positions and their colors. Figure 15(a) shows the result by interpolation, and Fig. 15(b) shows the actual shape and the quality of the material. In this case, the experimental result coincided with the actual one, too.

### C. Evaluation

The average error of 3-D measurement was about 1mm.

Of course, we can measure an object with curved surfaces like Fig. 16.

As to the hammering power, we strike the object in four conditions; strong, medium, weak, and very weak. Table I shows parameter $B$ and $V_f$ in these conditions. Parameter $B$ and $V_f$ does not change drastically when the hammering power changes, except for the power is very weak whose signal-to-noise ratio is very large. Therefore, it is verified that these parameters don't depend on the hammering power. This is because these parameters are less subject to the amplitude of the sound signal.

As to the computation time, image and sound signal processing need 2.3s per one hammering point on the average with a personal computer (CPU: Pentium IV 1.6GHz, Memory: 256MB, OS: Windows 2000). This shows that there is no problem for real-time processing because the interval of hammering is over 2s.

These results show the effectiveness of the proposed method.

## VI. CONCLUSION

In this paper, we propose a new method that can inspect visual and non-visual features of objects simultaneously
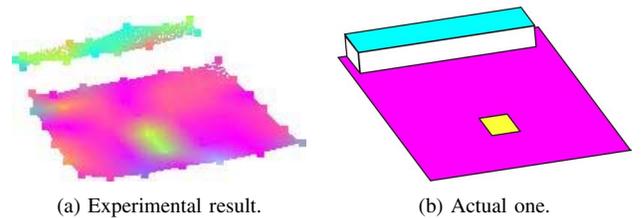

(a) Experimental result.
(b) Actual one.

Fig. 12. Result of hammering test for object I.

Plastic    Styrene foam

100
100                    100
Cork    100   100    Wood

(a) Schematic.          (b) Photo.

Fig. 13.   Object II.



Fig. 16.   Object with curved surfaces.



(a) Discrimination map.          (b) Hammering point.
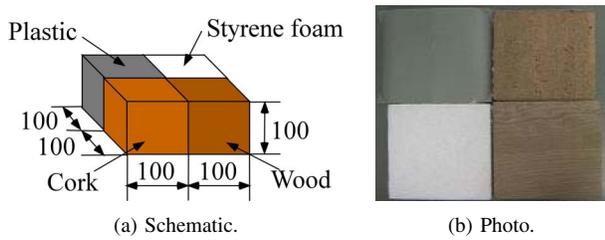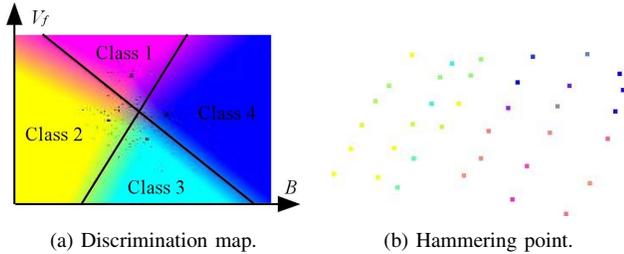
Fig. 14.  Discrimination map and 3-D positions and colors of each hammering point for object II.

hammering test over a wide area of an object.
• Display of the results is easy for human to understand by using a color.

As a future work, we must investigate various conditions concerning to shapes and materials of both an object and a hammer.

Another efficient feature should be investigated in order to make the method more accurate and robust. Speech analysis methods like Cepstrum analysis also can improve the robustness of sound signal processing.

The color values given for center positions of classes must be also selected according to the color balance of the discrimination map and applications of this method.

by using image and sound signal processing techniques. A hammering test system with a digital video camera and a personal computer is proposed. The system estimates nonvisual features by using the discrimination map that consists of the decay index and the frequency distribution from the sound signal processing, and determines simultaneously hammering positions by image signal processing. Validity of the method is confirmed by experiments. To sum up, our proposed method has the following advantages.

• It is not necessary to newly introduce special equipment to the hammering test.
• Our method can analyze the hammering sound quantitatively and objectively.
• The 3-D shape of the object whose surface has no texture can be measured, because the position and the orientation of the hammer are measured.
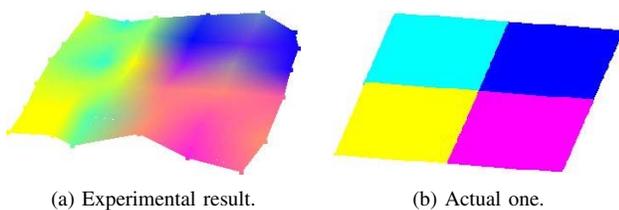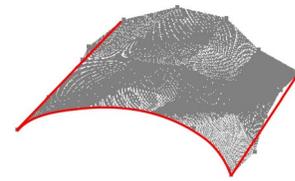• It is very easy to create database of results for the

REFERENCES

[1] M. J. Sansalon and W. B. Streett: *Impact-Echo*, Bullbrier Press, Ithaca, NY, 1997.
[2] C. Cheng and M. Sansalone: "The Impact-Echo Response of Concrete Plates Containing De-Laminations: Numerical, Experiential and Field Studies," *Materials and Structures*, Vo.26, pp.274–285, 1993.
[3] N. Kato, N. Saeki, Y. Tanigawa, K. Kato, T. Kawai and Y. Murata: "Evaluation of Deterioration and Specification of Defect of RC Structure by Natural Potential and External Impact-Acoustic Methods," *Materials Science Research International*, Vol.7, No.2, pp.138–146, 2001.
[4] T. Iwase, M. Doi, Y. Mahide and T. Abe: "Vibration Propagation Characteristics around Defective Part on Concrete Structure: Experiments on Diagnosis Based on Phase Delay in Vibration Response," *Acoustic Science and Technology*, Vol.22, pp.383–385, 2001.
[5] M. Asano, T. Kamada, M. Kunieda and K. Rokugo: "Frequency Characteristic of Impact-Acoustic Waves in Concrete Slab with Artificial Defects," *3rd Kumamoto International Workgroup on Fracture, Acoustic Emission and NDE in Concrete*, 2003.
[6] O. R. Gericke: "Determination of the Geometry of Hidden Defects by Ultrasonic Pulse Analysis Testing," *Journal of Acoustic Society of America*, Vol.35, pp.364–368, 1963.
[7] T. Yamaguchi, S. Hirono and T. Murobuchi: "Measurement of Crack Depth in RC Structures by Diffracted Wave Method," *Proceedings of International Conference NDT in Linepool*, pp.519–529, 1993.
[8] T. Hirata and T. Uomoto: "Detection of Ultrasonic Pulse Echo through Steel Bar in Concrete Crack Depth Measurement," *Non-Destructive Testing in Civil Engineering 2000*, pp.383–390, 2000.
[9] T. Kawai, M. Ito and N. Ikeda: "Damage Detection of Structure Using Image Processing (1st Report, Detection of Joint Failure)," *Transactions of the Japan Society of Mechanical Engineers* (*Series C*), Vol.67, No.653, pp.17–22, 2001, (in Japanese).
[10] G. H. Ball and D. J. Hall: *ISODATA – Novel Method of Data Analysis and Pattern Classification*, Stanford Research Institute, 1965.
[11] K. Takahashi, T. Sugiyama and K. Abe: "Region Segmentation of Color Images Based on ISODATA Clustering Method with Image Space Information," *Proceedings of the 4th Asian Conference on Computer Vision*, Vol.1, pp.216–221, 2000.
[12] B. K. P. Horn: *Robot Vision*, MIT Press, NY, 1986.
[13] R. Y. Tsai: "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses," *IEEE Journal of Robotics and Automation*, Vol.RA-3, No.4, pp.323–344, 1987.

(a) Experimental result.          (b) Actual one.

Fig. 15.   Result of hammering test for object II.

TABLE I
HAMMERING POWER AND PARAMETER $B$, $V_f$.

|        | Strong | Medium | Weak | Very Weak |
|--------|--------|--------|------|-----------|
| $B$    | 42.8   | 39.3   | 42.3 | 34.5      |
| $V_f$  | 183    | 169    | 167  | 159       |