Test and Evaluation of Japanese GPR-based AP Mine Detection Systems Mounted on Robotic Vehicles

This article introduces Japanese activities regarding a project, "Research and Development of Sensing Technology, Access and Control Technology to Support Humanitarian Demining of AP Mines." This project, which includes the research of six teams from academia and industry, has been funded by the Japan Science and Technology Agency (JST) under the auspices of the Ministry of Education, Culture, Sports, Science and Technology (MEXT). The developed systems are equipped with both ground-penetrating radar and a metal detector, and they are designed to make no explicit alarm and to leave decision-making of detection using subsurface images to the operators. To evaluate these kinds of systems, a series of trials was conducted in Japan from 8 February to 11 March 2005.

by Jun Ishikawa and Mitsuru Kiyota [Japan Science and Technology Agency] and Katsuhisa Furuta [Tokyo Denki University]

At current clearance speed, it will take more than 100 years to remove all the landmines that remain in the world.[1] Consequently, Japan is developing more efficient and safer humanitarian demining technologies. This article introduces Japanese robotic sensor systems that provide deminers with clear subsurface images via ground-penetrating radar in combination with metal detectors (GPR+MD).

Experiment Overview: Background

To reconstruct clear images, highly accurate sensor-positioning systems, as well as sensing technology itself, are indispensable because one of the most important pieces of information for signal processing is sensor position, where the sensor acquires a series of data for GPR+MD.

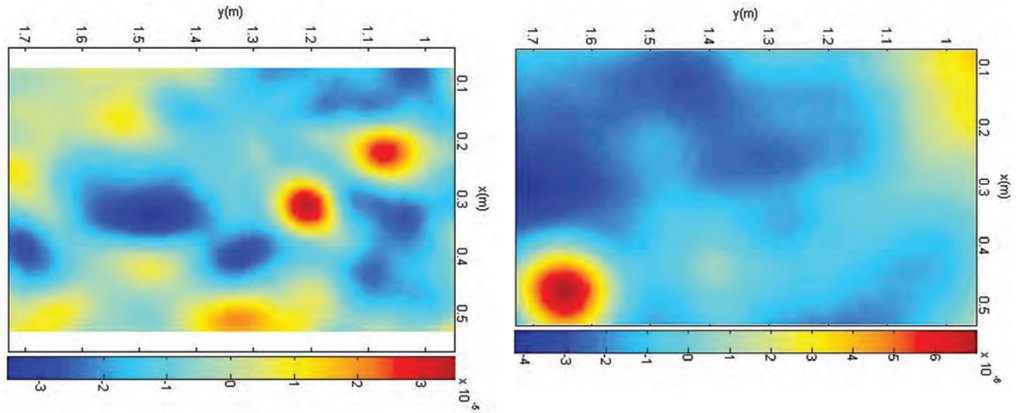
There are many kinds of anti-personnel landmines, which can be laid by humans or scattered by airplanes, and mined areas are not limited to plains but also marshes, canals, steep hillsides, seashores, deserts, mountains and forests. For such rough terrain, robotic systems must have sensor heads that can scan the ground as closely as possible but never touch it as well-trained deminers do. Metal detectors, which are a kind of an electromagnetic induction (EMI) sensor, have the possible detection distance of about 15 centimeters for minimum-metal landmines. For these metal detectors, it is a challenge for sensor systems to access minefields and manipulate the sensor head in severe environments in order to stay as close to the ground as possible. Thus, Japanese advanced robotics and sensor engineering have been fused to create novel detectors.

Japan started preparation for this kind of research and development in March 1997, when the Tokyo Conference on Anti-personnel Landmines was held. At this conference, participants undertook a comprehensive discussion to strengthen international efforts toward addressing the problems of AP landmines, especially landmine clearance by the United Nations and other organizations; development of new technology for mine detection and removal; and assistance to victims. In December 1997, Keizo Obuchi, then Minister for Foreign Affairs of Japan, signed the Ottawa Convention,[2] and the ultimate goal of zero victims was proposed. Since August 2002, the Japanese have undertaken preparations to start humanitarian-demining R&D.[3]

Japanese R&D of Anti-personnel Landmine Detection System

With strong expectations from the world community for Japanese contributions in this area, the Ministry of Education, Culture, Sports, Science and Technology established the Committee of Experts on Humanitarian Demining Technology in January 2002, believing in the importance of tackling the technological development of AP landmine detection using advanced Japanese technology. The Committee's findings were presented to MEXT in the report, "Promoting R&D for Humanitarian Demining Technology."[4] Based on this report, the Japan Science and Technology Agency announced a call for proposals for R&D projects in humanitarian-demining technology. Out of the 82 proposals, 12 projects were selected, and an R&D project named "Research and Development of Sensing Technology, Access and Control Technology to Support Humanitarian Demining of Anti-personnel Mines" started in October 2002.

The JST project is essentially divided into a short-term R&D project and a medium-term one. Because of the urgent need for this technology, the short-term R&D project is expected to have prototypes in field trials within three years. The JST medium-term R&D project is on a five-year schedule. The goal is to develop sensing technologies that can detect the explosive itself, in the range of about 30 to 100 grams.



All figures courtesy of Japan Science and Technology Agency.

Figure 1a: Detection images from stepped-frequency GPR. Horizontal slices showing two targets at a five-centimeter depth (left) and a target at a 25-centimeter depth (right).

Short-term R&D project. The objectives of the short-term R&D project are to develop sensing technology that can safely and efficiently detect AP landmines based on the physical differences between landmines and soils, and to develop access devices and manipulation technology that carry sensors into minefields and allow them to scan the ground precisely. More specifically, the goal is to develop vehicle-mounted GPR+MD dual-sensor systems that make no explicit alarm and provide operators with clear subsurface images. This means that the decision to determine whether or not a shadow in the image is a real AP landmine is entirely left to the operator, similar to how medical doctors can find cancer by reading CT images. This feature discriminates the systems from conventional GPR+MD dual sensors that are based on alarm tones.

In the short-term project, four sensors and three robotic vehicles have been developed. One of those is the Mine Hunter Vehicle. The vehicle itself and the manipulator have been developed by a research team of Professor Kenzo Nonami's at Chiba University.[5] The MHV can interchangeably mount two GPR sensors in addition to a commercial, off-the-shelf metal detector.

One sensor is a stepped-frequency GPR developed by Professor Motoyuki Sato's team at Tohoku University,[6] hereinafter referred to as MHV #1. Stepped-frequency radar determines distance to a target by constructing a synthetic range profile, which is a time domain approximation derived from the frequency response of a combination of stepped-frequency signals via inverse fast Fourier transform (IFFT). The major advantage of the stepped-frequency method is that the spectrum bandwidth can be easily tuned to fit an optimum value according to environmental conditions such as soil moisture.

The other sensor is an impulse GPR developed by Professor Ikuo Arai's project at the University of Electro-Communications,[7] hereinafter referred to as MHV #2. This kind of GPR operates by transmitting a very narrow pulse (< 1 nanosecond) of electromagnetic wave, the advantage of which is that the measurement time required to generate one range profile is very short. After the GPR scans to acquire a range profile for every interval of several centimeters,[8] GPR tomography gives subsurface horizontal slices as shown in Figure 1a, and further calculation provides operators with three-dimensional images (Figure 1b).

Professor Toshio Fukuda's group at Nagoya University developed a dual sensor with built-in stepped-frequency GPR+MD.[9]

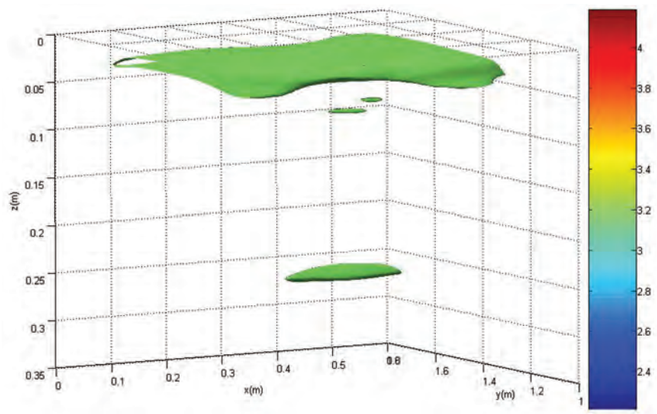


Figure 1b: Detection images from stepped-frequency GPR. Three-dimensional image of three targets in the horizontal slices.

The sensor system scans the ground, being carried by a low-reaction-force manipulation frame that has four balloons on the legs to softly land it on minefields. The manipulation frame is attached to the top of a boom of a crane vehicle developed by Mr. Tomohiro Ikegami's group at TADANO Ltd. The vehicle has a 20-meter reach for a 200-kilogram payload with a positioning accuracy of 15 centimeters. These elements have been integrated into the Advanced Mine Sweeper (AMS), which can adapt to various geographical environments.[10]

Professor Shigeo Hirose's team at the Tokyo Institute of Technology developed the Gryphon buggy system, which can be remotely controlled to access minefields.[11] The manipulator mounted on the buggy has been designed to cancel reaction force induced by sensor scanning.[12] The sensor is a GPR+MD dual sensor named the Advanced Landmine Imaging System (ALIS), and it can also be used as a handheld detector.[13] ALIS was developed by Professor Sato's team and underwent a field trial in Afghanistan in December 2004.

Medium-term R&D project. Professor Hideo Itozaki's group of Osaka University is developing a nuclear quadrupole-resonance detector.[14] In the analysis, a radio-frequency electromagnetic wave is first emitted and excites nuclear spin of ^{14}N in explosives. Then a magnetic wave detector, such as an induction coil, detects subsequent NQR signals from the ^{14}N if any intended target exists, and the resonance frequency of the signal is unique for each explosive material. Thus explosives can be identified.

Two research teams on the project are trying to develop detectors based on the neutron analysis identifying explosives through back-scattering of neutrons and detection of specific energy gamma rays

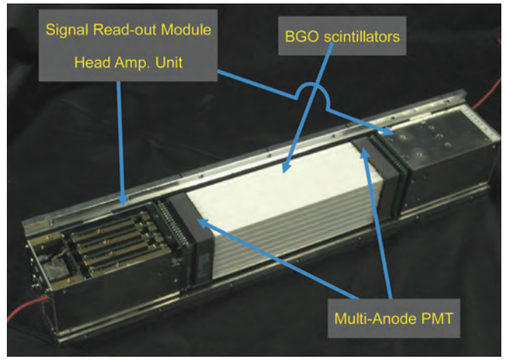


Figure 2: Multi-Compton gamma camera based on stacked BGO scintillator rods.

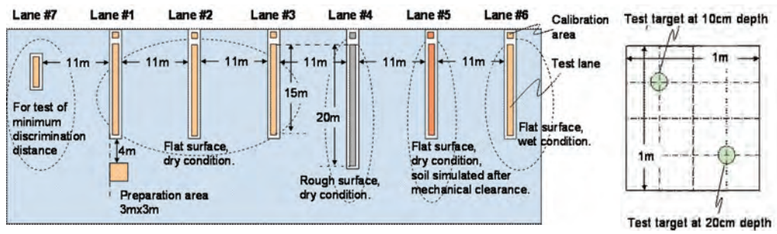


Figure 3: Test-lane layout and the calibration area.

from capture on hydrogen and nitrogen atoms of explosives. Professor Kiyoshi Yoshikawa's group from Kyoto University has prototyped an extremely compact neutron source based on an inertial-electrostatic confinement fusion device 20 centimeters in diameter.[15] Professor Tetsuo Iguchi's group of Nagoya University has prototyped another neutron source, which is an improved Cockcroft-Walton-type accelerator neutron source using a deuterium-deuterium (DD) fusion reaction. They have also developed a prototype of a multi-Compton gamma camera, which estimates the incoming direction of 10.8MeV gamma-rays produced from the nitrogen of the explosive (Figure 2).[16]

The medium-term R&D project is expected to have prototypes in field trials within five years, namely in 2007, in combination with one of the prototypes of MHV, AMS or Gryphon.

Experimental Design[17,18]

To evaluate the short-term R&D prototypes, a series of tests was conducted from 8 February to 11 March 2005 in Sakaide City, Japan. Seven test lanes were constructed using more than 200 landmine surrogates (Figure 4). Since operators' pre-knowledge of the locations of buried targets significantly influences the detection results for such systems that make no explicit alarm, lanes 1 to 6 are designed to be used for blind tests.

Test lanes and landmine surrogates. In constructing test lanes, all the original soil was removed from a width of 2 meters to a depth of 0.5 meters in the vertical section, and the lanes were filled with homogeneous and non-mineralized (cooperative) soil. The actual width of test lanes is 1 meter, and mine surrogates were buried shallower than or equal to a depth of 0.3 meters (1 foot). The features of each lane are as follows:

- Lanes 1, 2 and 3 are 15 meters long with a flat surface.
- Lane 4 is 20 meters long, with 15 bumps in the surface, each with a height of 10 centimeters and a diameter of 60 centimeters, and small stones are mixed to make the soil heterogeneous.
- Lane 5 simulates minefields in post-clearance inspection after mechanical demining, with the soil stirred and not packed.
- Lane 6 is wet, with 10 liters of water per square meter sprinkled one hour before the test starts.

Figure 4 shows four kinds of landmine surrogates used in the test. The M14[19] and PMN2[20] contain a metal part—an 18-millimeter[21] vertical carbon steel pin with a diameter of 3 millimeters—and the Type72[22] has a 4-millimeter vertical carbon steel pin with a diameter of 4 millimeters. The Type72-S[23] mine is made by modifying a product of Amtech Aeronautical Limited[24] and has exactly the same metal part as the International Test Operations Procedure standard I_0, a 12.7-millimeter vertical aluminum tube. Silicone rubber was substituted for explosives in all the surrogates.

Experimental design. Through the tests, influences of various factors on probability of detection should be evaluated. Namely, in Experiment 1, target types, target depth, soil conditions and target angles were chosen as factors to be tested. There are two or four levels for each factor as described in Table 1. According to the soil conditions, for example, targets (landmine surrogates) that are classified into "flat," "wet," "stirred" and "rough" are respectively buried in lanes 1–3, 6, 5 and 4 at a specified depth and angle as defined in Figure 5. Experiment 2 was designed to mainly evaluate

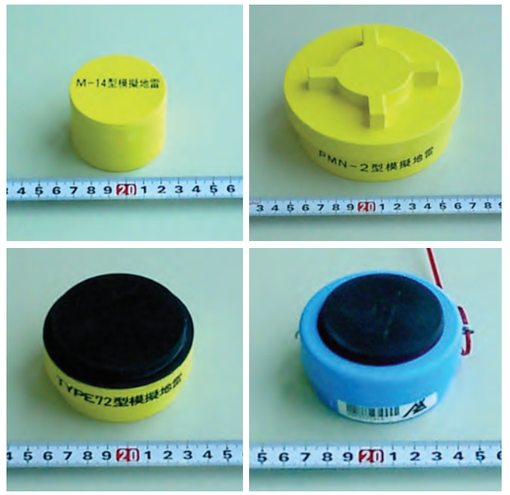


Figure 4: Landmine surrogates used in test.

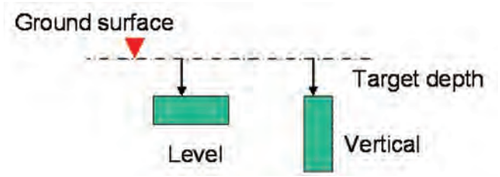


Figure 5: Definitions of target depth and angle.

the minimum discrimination distance. Two levels were chosen for the factor "distance to adjacent target" as described in Table 2. One level consists of pairs of targets in a distance of 15 centimeters and the other level consists of independent targets, the separation of which shall be at least 50 centimeters.

Due to the limitation of time for the trial and the number of targets, it is impossible to test all the combinations of levels in Tables 1 and 2. To impartially collect unbiased data for statistical analysis under this limitation, orthogonal experimental designs based on $L_{16}(2^{15})$ and $L_8(2^7)$ orthogonal arrays were respectively used for Experiments 1 and 2. Assigning the columns of the array to each factor as specified in Tables 1 and 2 derives a reduced set of combinations, the results of which are summarized in Tables 3 and 4. For example, the number of experimental runs can be reduced from 128 (4×4×4×4×2) to 16 in Experiment 1.

According to Tables 3 and 4 (see opposite page), all the targets were buried at random locations in the specified lanes and were left for more than one month before the test began. Testees can submit all the impartial data needed for statistical analysis by reporting detection results from lanes 1 through 6. In the trial, at least two testees from every device took the test in all 6 lanes.

Factor	Number of Level	1	2	3	4	Assigned column
A: Target type	4	M14	PMN2	TYPE72	TYP72-S	4,8,12
B: Target depth	4	0cm	10cm	20cm	30cm	5,10,15
C: Soil condition	4	Flat	Wet	Stirred	Rough	7,9,14
D: Target angle	2	Vertical	Level			6

Table 1: Factors A to D and the levels for Experiment 1.

Factor	Number of Level	1	2	3	4	Assigned column
A: Distance to adjacent target	2	15cm	> 50cm			1
B: Target depth	4	0cm	10cm	20cm	30cm	2,4,6
C: Soil condition	2	Flat	Rough			7

Table 2: Factors A to C and the levels for Experiment 2.

Benchmarking. To compare performance of the GPR+EMI dual sensors with that of existing metal detectors, a benchmarking trial was conducted. Namely, a tester who knew the exact positions of targets checked if any metal-detector response occurred just above every buried target. The result of this test shows the best performance of the metal detectors used.

Test procedures. Testees took blind tests for each lane following the procedures as described below:

1. Before the test starts, the tester records temperature, relative humidity and volumetric water content that is measured by time domain reflectometry (TDR).[25]
2. The testee does close-in detection work using a sensor system cooperatively with vehicle operators.
3. After the work finishes, the tester records temperature, relative humidity and volumetric water content measured by TDR.
4. The testee reports the following data for every detected anomaly:
   - Coordinates of the detected target
   - Depth of the detected target
   - Confidence rating defined in Table 5 and the final decision whether or not to declare the anomaly as a landmine surrogate
5. The tester determines whether the declared anomaly can be considered to be from the intended targets,[26] that is, within a detection halo, the radius of which is half of the target diameter plus 10 centimeters.[27]
6. Finally, the tester classifies the reported data into four categories:
   - **True positive:** The testee declared it as a target and this is true.
   - **False positive:** The testee declared it as a target and this is not true. This is a false alarm.
   - **True negative:** The testee declared it as a fragment, clutter or noise and this is true.
   - **False negative:** The testee declared it as a fragment, clutter or noise and this is not true. This is missing a target.

Completing the tests from lanes 1 through 6 means that the testee finished all 24 experimental runs of Experiments 1 and 2 described in Tables 3 and 4.

The most important thing is to practically use these technologies to improve landmine-detection efficiency and reduce minefields. To do so, the mine-detection systems must be robust, simple and highly cost-effective. The Japanese domestic trial is the first step.

Test and Evaluation Results

The following is the data analysis and evaluation of test results for anti-personnel landmine detection systems using ground-penetrating radar mounted on robotic vehicles for humanitarian demining.[17,18]

The test results showed that combining GPR with metal detectors can improve probability of detection for targets around a depth of 20 centimeters, where it is difficult to detect the targets by using only a metal detector. It has also been learned that positioning control must be improved in scanning the ground with a sensor head, which is key to making the best of use of metal detectors mounted on vehicles. Lessons learned have been reflected in further improvement of the prototypes. In the following sections, data analysis, methods and evaluation results are described.

Data analysis. According to the experimental design proposed above, data from eight testees (two each from every system) have been acquired. The comprehensive results of probability of detection (PD) are shown in Tables 6 and 7 and were acquired through Experiments 1 and 2. The systems named are anonymous and described as Device

Continued on page 98, TEST

Experimental run	Target type	Target depth	Soil condition	Target angle	Number of target	Lane used in test
1	M14	0cm	Flat	Vertical	7	Lane #3
2	PMN2	10cm	Wet	Vertical	7	Lane #6
3	Type72	20cm	Stirred	Level	7	Lane #5
4	Type72-S	30cm	Rough	Level	7	Lane #4
5	M14	10cm	Stirred	Level	7	Lane #5
6	PMN2	0cm	Rough	Level	7	Lane #4
7	Type72	30cm	Flat	Vertical	7	Lane #3
8	Type72-S	20cm	Wet	Vertical	7	Lane #6
9	M14	20cm	Rough	Vertical	7	Lane #4
10	PMN2	30cm	Stirred	Vertical	7	Lane #5
11	Type72	0cm	Wet	Level	7	Lane #6
12	Type72-S	10cm	Flat	Level	7	Lane #3
13	M14	30cm	Wet	Level	7	Lane #6
14	PMN2	20cm	Flat	Level	7	Lane #2
15	Type72	10cm	Rough	Vertical	7	Lane #4
16	Type72-S	0cm	Stirred	Vertical	7	Lane #5

Table 3: Design result for Experiment 1.

Experimental run	Distance to adjacent target	Target depth	Soil condition	Number of target	Lane used in test
1	15cm	0cm	Flat	14	Lane #2
2	15cm	10cm	Rough	14	Lane #4
3	15cm	20cm	Rough	14	Lane #4
4	15cm	30cm	Flat	14	Lane #1
5	> 50cm	0cm	Rough	14	Lane #4
6	> 50cm	10cm	Flat	14	Lane #3
7	> 50cm	20cm	Flat	14	Lane #1
8	> 50cm	30cm	Rough	14	Lane #4

Table 4: Design result for Experiment 2.

0	25	50	75	100
I'm 100% sure that there is nothing here.	It seems that there is something here.	I'm 100% sure that there is something here.	I'm not totally sure, but I would say the detected object seems to be a landmine.	I'm 100% confident that the detected object is a landmine.

Table 5: Definition of confidence rating.

Experimental run	Target type	Target depth	Soil condition	Target angle	Lane used	Metal Detector	Device #1		Device #2		Device #3		Device #4	
							Testee #1	Testee #2	Testee #3	Testee #4	Testee #5	Testee #6	Testee #7	Testee #8
1	M14	0cm	Flat	Vertical	Lane #3	1.000	1.000	0.857	1.000	1.000	1.000	1.000	1.000	0.714
2	PMN2	10cm	Wet	Vertical	Lane #6	1.000	0.571	0.571	0.857	1.000	0.429	0.429	1.000	0.571
3	Type72	20cm	Stirred	Level	Lane #5	0.000	0.571	0.286	0.143	0.143	0.000	0.429	0.857	0.429
4	Type72-S	30cm	Rough	Level	Lane #4	0.143	0.286	0.286	0.286	0.143	0.143	0.143	0.714	0.286
5	M14	10cm	Stirred	Level	Lane #5	1.000	0.857	0.714	1.000	1.000	1.000	1.000	0.857	0.286
6	PMN2	0cm	Rough	Level	Lane #4	1.000	1.000	0.857	1.000	1.000	0.286	0.429	0.857	0.571
7	Type72	30cm	Flat	Vertical	Lane #3	0.143	0.714	0.286	0.143	0.000	0.286	0.286	1.000	0.143
8	Type72-S	20cm	Wet	Vertical	Lane #6	0.143	0.429	0.714	0.143	0.143	0.429	0.429	0.857	0.000
9	M14	20cm	Rough	Vertical	Lane #4	0.143	0.286	0.286	0.000	0.000	0.000	0.000	0.429	0.143
10	PMN2	30cm	Stirred	Vertical	Lane #5	0.000	1.000	0.857	0.429	0.000	0.429	0.286	0.857	0.000
11	Type72	0cm	Wet	Level	Lane #6	1.000	0.857	1.000	1.000	1.000	0.714	0.857	1.000	0.857
12	Type72-S	10cm	Flat	Level	Lane #3	1.000	0.857	1.000	1.000	1.000	0.857	0.857	1.000	0.571
13	M14	30cm	Wet	Level	Lane #6	0.000	0.143	0.286	0.143	0.000	0.000	0.143	0.429	0.429
14	PMN2	20cm	Flat	Level	Lane #2	0.571	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.429
15	Type72	10cm	Rough	Vertical	Lane #4	0.286	0.286	0.429	0.143	0.286	0.286	0.143	0.714	0.000
16	Type72-S	0cm	Stirred	Vertical	Lane #5	1.000	0.857	1.000	1.000	1.000	1.000	0.857	0.857	0.429

Table 6: PD of eight testees of Experiment 1. Highlighted data of four testees are analyzed as shown in Figure 13.

TEST, Continued from page 96

1, 2, 3 and 4. A benchmarking result is also shown in the tables. This section discusses how the data are analyzed.

Analysis of variance (ANOVA). ANOVA tests are necessary if there are significant differences of PD between levels for each factor.[28] This is useful to check if experiments are well-designed to discuss influences of the factors on PD and to see how the factors interfere in PD. Some levels such as a target depth of 30 centimeters have been set to be very difficult in comparison with the sensor specifications because an objective of the test is to make the limitations of the sensor systems clear.

In the following part of this section, an example is given for an ANOVA of Experiment 2 assuming that an experimental result in Table 8 is acquired from a system with no repetition. First the mean of the results is calculated as:

$$\mu = \frac{1}{8} \cdot \sum_{i=1}^{8} p_i \qquad \textit{Equation 1}$$

and the main effect for each level of the factors A, B and C is derived as follows:

$$a_{15cm} = \frac{1}{4} \cdot (p_1 + p_2 + p_3 + p_4) - \mu$$

$$a_{MT\,50cm} = \frac{1}{4} \cdot (p_5 + p_6 + p_7 + p_8) - \mu \qquad \textit{Equation 2}$$

$$b_{0cm} = \frac{1}{2} \cdot (p_1 + p_5) - \mu$$

$$b_{10cm} = \frac{1}{2} \cdot (p_2 + p_6) - \mu$$

$$b_{20cm} = \frac{1}{2} \cdot (p_3 + p_7) - \mu$$

$$b_{30cm} = \frac{1}{2} \cdot (p_4 + p_8) - \mu \qquad \textit{Equation 3}$$

$$c_{Flat} = \frac{1}{4} \cdot (p_1 + p_4 + p_6 + p_7) - \mu$$

$$c_{Rough} = \frac{1}{4} \cdot (p_2 + p_3 + p_5 + p_8) - \mu \qquad \textit{Equation 4}$$

Next, error effects e_i for i = 1, L, 8 are calculated as:

$$e_1 = p_1 - (\mu + a_{15cm} + b_{0cm} + c_{Flat})$$
$$e_2 = p_2 - (\mu + a_{15cm} + b_{10cm} + c_{Rough})$$
$$e_3 = p_3 - (\mu + a_{15cm} + b_{20cm} + c_{Rough})$$
$$e_4 = p_4 - (\mu + a_{15cm} + b_{30cm} + c_{Flat})$$
$$e_5 = p_5 - (\mu + a_{MT\,50cm} + b_{0cm} + c_{Rough})$$
$$e_6 = p_6 - (\mu + a_{MT\,50cm} + b_{10cm} + c_{Flat})$$
$$e_7 = p_7 - (\mu + a_{MT\,50cm} + b_{20cm} + c_{Flat})$$
$$e_8 = p_8 - (\mu + a_{MT\,50cm} + b_{30cm} + c_{Rough}) \qquad \textit{Equation 5}$$

Experimental run	Distance to adjacent target	Target depth	Soil condition	Lane used	Metal Detector	Device #1		Device #2		Device #3		Device #4	
						Testee #1	Testee #2	Testee #3	Testee #4	Testee #5	Testee #6	Testee #7	Testee #8
1	15cm	0cm	Flat	Lane #2	1.000	0.929	0.857	0.929	0.929	0.786	0.786	1.000	1.000
2	15cm	10cm	Rough	Lane #4	0.500	0.643	0.786	0.786	0.500	0.429	0.214	0.929	0.286
3	15cm	20cm	Rough	Lane #4	0.286	0.500	0.571	0.214	0.071	0.071	0.143	0.857	0.571
4	15cm	30cm	Flat	Lane #1	0.071	0.500	0.786	0.143	0.071	0.286	0.286	0.714	0.429
5	> 50cm	0cm	Rough	Lane #4	1.000	0.714	0.857	0.929	0.929	0.214	0.214	0.857	0.643
6	> 50cm	10cm	Flat	Lane #3	1.000	0.786	0.929	1.000	1.000	0.929	0.857	1.000	0.643
7	> 50cm	20cm	Flat	Lane #1	0.071	0.714	0.786	0.857	0.714	0.500	0.714	1.000	0.214
8	> 50cm	30cm	Rough	Lane #4	0.143	0.214	0.357	0.143	0.071	0.143	0.143	0.571	0.143

Table 7: PD of eight testees of Experiment 2. Highlighted data of four testees are analyzed as shown in Figure 14.

Experimental run	Distance to adjacent target	Target depth	Soil condition	Experimental results (PD)
1	15cm	0cm	Flat	p_1
2	15cm	10cm	Rough	p_2
3	15cm	20cm	Rough	p_3
4	15cm	30cm	Flat	p_4
5	> 50cm	0cm	Rough	p_5
6	> 50cm	10cm	Flat	p_6
7	> 50cm	20cm	Flat	p_7
8	> 50cm	30cm	Rough	p_8

Table 8: Notion of detection of probability for ANOVA example.

Now, for example, a linear model for the probability of detection $p1$ can be defined as:

$$p1 = \mu + a15cm + b0cm + cFlat + e1 \qquad \textit{Equation 6}$$

For the ANOVA, four means of squares (variances) are calculated as follows:

$$V_A = \frac{S_A}{f_A} = 4 \cdot \left[a_{15cm}{}^2 + a_{MT\,50cm}{}^2 \right] / f_A \qquad \textit{Equation 7}$$

$$V_B = \frac{S_B}{f_B} = 2 \cdot \left[b_{0cm}{}^2 + b_{10cm}{}^2 + b_{20cm}{}^2 + b_{30cm}{}^2 \right] / f_B \qquad \textit{Equation 8}$$

$$V_C = \frac{S_C}{f_C} = 4 \cdot \left[c_{Flat}{}^2 + c_{Rough}{}^2 \right] / f_C \qquad \textit{Equation 9}$$

$$V_e = \frac{S_e}{f_e} = \left[\sum_{i=1}^{8} e_i{}^2 \right] / f_e \qquad \textit{Equation 10}$$

where f_C and f_e are the degrees of freedom of factors and error.

By comparing the variances due to levels of each factor (i.e., V_A, V_B and V_C with the variance due to measurement error (V_e) using F-test), the significance of the differences between levels is tested. In this test, the null hypothesis is that the main effects of levels for a factor are all equal (i.e., there is no difference in influences of levels for the factor on PD). The computed F statistic in Table 9 follows an F distribution with corresponding degrees of freedom under the assumption that variances of PD have homogeneity.[29] Therefore, the significance of F can be determined in the usual way by using the table of F. If the computed value of F is larger than the tabled value, the null hypothesis is rejected. This means that at least one pair of main effects is significantly different.

As described above, detection results reported by testees are classified into four categories: true positive, false positive, true negative and false negative. However, the classification based on a testee's discrimination threshold is a one-sided view, and the number of true positives and the number of false positives change as the threshold is varied. An ROC curve shows us the relationship between the true positive and false positive for a variety of different thresholds, thus helping the determination of an optimal threshold as well as the comparison of sensor performance.

To plot an ROC curve, two histograms, which are measured on an interval scale in the confidence rating reported by the testee, are needed. One is from signals of intended targets that consist of true positives and false negatives, and the other is from signals of fragments, clutters or noise (i.e., true negatives and false positives). According to the histograms, the ratio of true positive (i.e., probability of detection) is plotted as a function of the ratio of false positive at every confidence

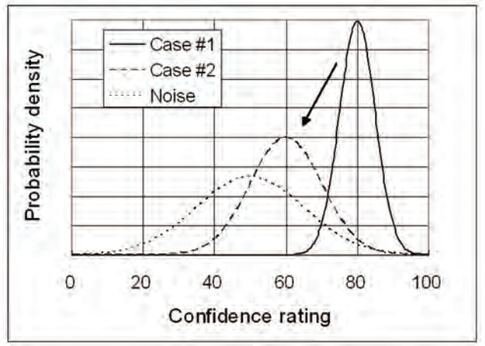


Figure 6: Normalized histogram of signal and noise.

Source of Variation	Degree of freedom	Sum of squares	Mean of squares	Computed F statistic
A: Distance to adjacent target	$f_A = 1$	S_A	$V_A = S_A / f_A$	$F_A = V_A / V_e$
B: Target depth	$f_B = 3$	S_B	$V_B = S_B / f_B$	$F_B = V_B / V_e$
C: Soil condition	$f_C = 1$	S_C	$V_C = S_C / f_C$	$F_C = V_C / V_e$
E: error	$f_e = 2$	S_e	$V_e = S_e / f_e$	

Table 9: Analysis of variance (ANOVA).

The 95-percent confidence limit of each main effect is experimentally derived by using V_e, the mean of squares due to error. For example, the 95-percent confidence interval of a_{15cm} is given by:

$$a_{15cm} \pm t_{f_e,\,95\%} \cdot \sqrt{\frac{V_e \cdot (f_A + 1)}{n_d}} \qquad \textit{Equation 11}$$

where n_d = 8 is the total number of experiments (the number of experimental runs multiplied by repetitions), and $t_{fe',95\%}$ is the quantile of the t-distribution for probability 95 percent with f_e degrees of freedom.

Receiver operating characteristic curve. It has been 30 years since radiographic applications of ROC curves were reported[30] and it is well-known that analysis based on ROC curves is suitable for subjective evaluation of imaging equipment. In the test and evaluation here, ROC curves were also used to evaluate sensor effectiveness in terms of both PD and false-alarm rate.

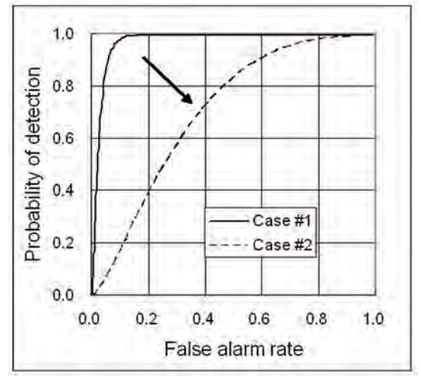


Figure 7: Example of ROC curves.

rating (threshold). As shown in Figure 6, if a sensor functions well, a histogram of targets (solid line) is distributed apart from that of noise (dotted line), and the resulting ROC curve climbs rapidly toward the upper left-hand corner of the graph as shown by the solid line in Figure 7. On the other hand, if another sensor gives a histogram of targets that is distributed closer to that of noise, the resulting ROC curve gets closer to a diagonal line as shown by the dashed line in Figure 7. This means the discriminating power decreases. Once ROC curves are obtained, there are many methods to test the difference between ROC curves.[31]

In the experiment, the number of true positives is controlled, but the number of false positives depends on how many false alarms are reported by the testee. Therefore, all the histograms discussed here are normalized by dividing frequencies by the total number of the population.

Experimental Results

Figure 8 shows the ground truth of the lane 2, and Figures 9 and 10 shows subsurface images from a sensor system. In this case, it has been shown that a metal detector can clearly image seven pairs of Type72 surrogates buried flush (Figure 9), and that a GPR sensor can display seven PMN2 surrogates at a depth of 20 centimeters (8 inches) (Figure 10), where the metal detector was not able to get any signal. Based on these kinds of images, testees have derived their detection results, and this section discusses the experimental results.

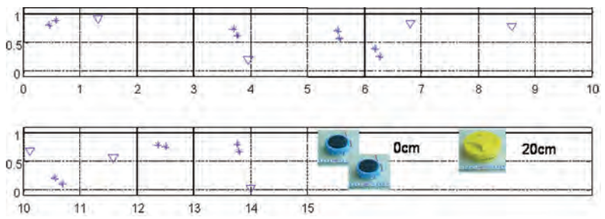


*Figure 8: Ground truth of the lane 2; ** shows a pair of Type72 and ▽ shows PMN2.*

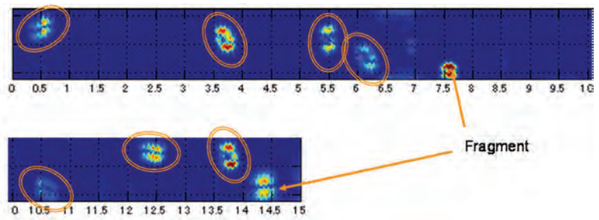


Figure 9: Detection image from a metal detector.

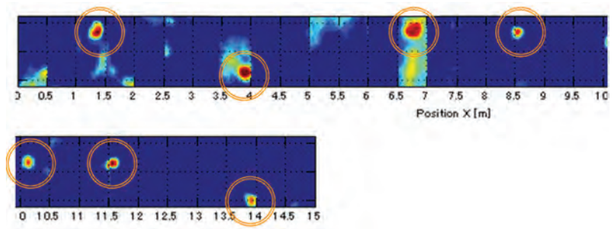


Figure 10: Detection image from a GPR sensor.

Probability of detection. The number of testees is eight, the breakdown of which is two from MHV with a step-frequency GPR+MD (MHV #1), two from MHV with a pulse GPR+MD (MHV #2), two from the Advanced Mine Sweeper with a step-frequency GPR+MD, and two from Gryphon with a pulse GPR+MD. The eight sets of data were analyzed by ANOVA to see the effects of factors. Note that the order of the systems is not consistent with devices 1–4 to keep anonymity.

Tables 10 and 11 show ANOVA results for Experiments 1 and 2, respectively, and Figures 11 and 12 show plots of factor effects (i.e., main effects added to the mean μ with 95-percent confidence intervals derived in the same way as Equation 11). In Tables 10 and 11, factors, the null hypothesis of which has been rejected at the level of significance of 0.05/0.01, are indicated by * (0.05) /** (0.01). For those factors, there have been significant differences in PD between the levels, and it can be said that it is meaningful to discuss how those factors influence PD and that the test lanes were well-designed to evaluate the sensor systems. It has been shown that there is a strong dependence of PD on target depth and that the developed systems still have problems for rough and uneven ground surface (Figures 11 and 12). Regarding factor A of Experiment 2, distance to adjacent target, the ANOVA showed that there was no significant difference in PD between a pair of Type72-S surrogates at a 15-centimeter distance and the other independent Type72-S surrogates.

Averages of PD of four testees, that is, one each from every system, are plotted in Figures 13 and 14, compared with the benchmarking result using only a metal detector. Confidence limits can be calculated the way that K. M. Simonson discusses in the *Sandia Report*[32,33] as the number of population for each level is derived from Tables 10 and 11 above. These results showed that the PD for targets deeper than 10 centimeters can be improved by combining GPR with MD. On the other hand, as also shown in Figures 13 and 14, some of the GPR+MD results in shallow levels were worse than those of metal detectors. This is because sensor height above the ground, which is controlled by manipulators, is higher than that of manual scanning of metal detectors, and this is considered to be improved by modifying the manipulation algorithm of a robotic part.

Lessons learned. Through the test and evaluation process, many lessons have been learned, some of which are listed below:

- The provided calibration area should have contained landmine surrogates for all levels of factors. Coaching a typical image for each level would much improve the detection rate.
- In some cases (for example, like Testee 7), high PDs have been accompanied by high false-alarm rates around 30 times/square meter,[34] and it was also proven that confirming the source of false alarms for GPR is much more difficult than those of metal detectors (i.e., metal fragments). Therefore, another performance index to penalize these GPR false alarms will be needed.
- PD in deep levels of 20–30 centimeters can be improved by combining GPR with MDs.

Continued on page 102, TEST

Source of Variation	Degree of freedom	Sum of squares	Mean of squares	Computed F statistic	Probability at the F value
A: Target type	3	0.9814	0.3271	5.6203	0.0012**
B: Target depth	3	5.4747	1.8249	31.3557	0.0000**
C: Soil condition	3	2.9087	0.9696	16.6598	0.0000**
D: Target angle	1	0.5000	0.5000	8.5911	0.0041**
e: error	117	6.8122	0.0582	-	
Total	127	16.6770			

Table 10: Eight testees' result of ANOVA of Experiment 1.

Source of Variation	Degree of freedom	Sum of squares	Mean of squares	Computed F statistic	Probability at the F value
A: Distance to adjacent target	1	0.0498	0.0498	1.0020	0.3210
B: Target depth	3	2.2235	0.7412	14.9135	0.0000**
C: Soil condition	1	1.0913	1.0913	21.9577	0.0000**
e: error	58	2.8827	0.0497	-	
Total	63	6.2472			

Table 11: Eight testees' result of ANOVA of Experiment 2.

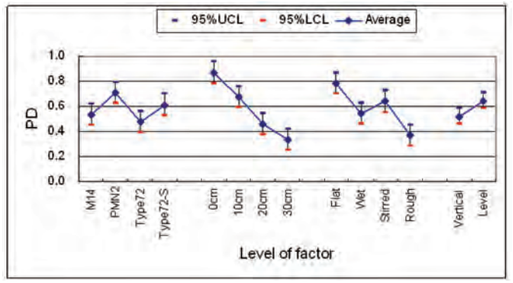


Figure 11: Factor effects of Experiment 1 with 95% confidence intervals

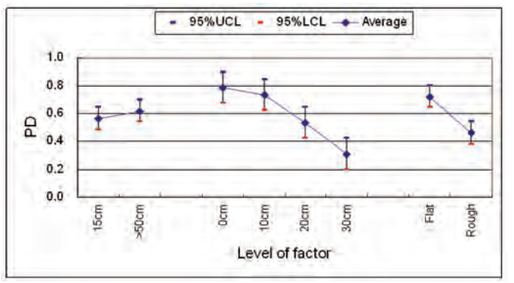


Figure 12: Factor effects of Experiment 2 with 95% confidence intervals.

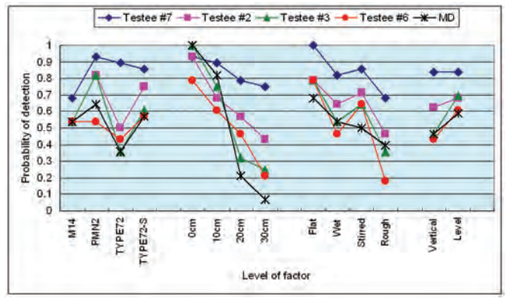


Figure 13: Averages of PD for Experiment 1. Testees 7, 2, 3 and 6 were chosen from each device.

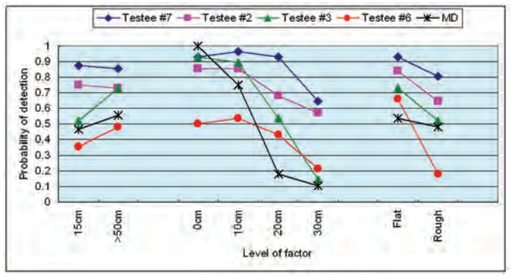


Figure 14: Averages of PD for Experiment 2. Testees 7, 2, 3 and 6 were chosen from each device.

Lane #	Device #1		Device #2		Device #3		Device #4	
	Testee #1	Testee #2	Testee #3	Testee #4	Testee #5	Testee #6	Testee #7	Testee #8
1	11.3	12.4	1.3	1.6	2.7	2.2	20.9	6.0
2	8.5	6.6	1.7	0.7	1.1	1.9	35.0	7.5
3	9.3	8.3	3.2	1.0	2.4	2.1	52.5	6.4
4	15.4	16.7	4.2	1.3	3.9	3.5	36.9	4.6
5	16.0	9.5	0.5	0.7	6.0	2.5	31.9	8.9
6	9.5	12.3	0.9	1.7	4.5	1.7	20.6	8.5

Table 12: False-alarm rate (1/square meter) of eight testees for each lane.

TEST, Continued from page 100

Evaluation of FAR. As described above, ROC curves are useful to see the qualification of sensors, taking into account tradeoff between PD and false alarm rate. Table 12 shows the FAR of eight testees for each of the six lanes in the experiment.

Figures 15a through 15d show typical ROC curves of testees 7 and 3 for lanes 2 and 4. Lane 2 has 21 targets buried as shown in Figure 8 (see page 102), and lane 4 with rough ground surface has 77 targets. A horizontal axis of each plot shows the normalized FAR, and the number of false alarms can be derived by FAR multiplied by the total number of negatives that is shown in each plot. In the case of Figure 15a, 65 percent of targets were detected with 100-percent confidence, but the other targets got mixed in 525 negatives. In Figure 15b, 95 percent of the targets were detected with 100-percent confidence. Figure 15c for lane 4 shows that testee 7 could not discriminate the targets from 738 negatives although the PD was 77 percent. On the other hand, as shown in Figure 15d, testee 3, the PD of which was 50 percent, detected 40 percent of targets with 100-percent confidence. These kinds of data have been used to optimize the operator's decision threshold and sensor sensitivities, and to improve the sensor performance.

Conclusions

Through the test and evaluation, many lessons have been learned, and these results were fed back to the testees for further improvement. The next step of the project is field trials in some mine-affected countries to confirm the improvements and to evaluate robustness and cost-effectiveness.

The authors would like to thank all the project members, especially the principal partners for the trial: Tohoku University, Chiba University, Tokyo Institute of Technology, University of Electro-Communications, Nagoya University, Kyoto University, Osaka University, TADANO Ltd., Mitsui Engineering & Shipbuilding Co., Fuji Heavy Industries Ltd., TAU GIKEN Co. Ltd. and Tokyo Gas Co.

See Endnotes, page 112

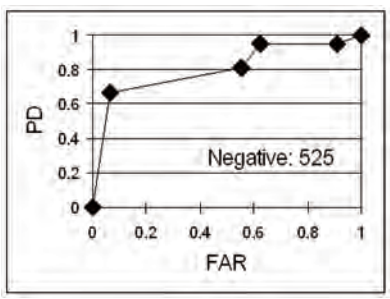


Figure 15a: ROC curve for lane 2 (testee 7). The total number of negatives (fragments, clutters or noise) is shown.

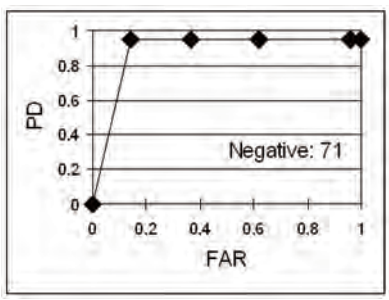


Figure 15b: ROC curve for lane 2 (testee 3). The total number of negatives (fragments, clutters or noise) is shown.

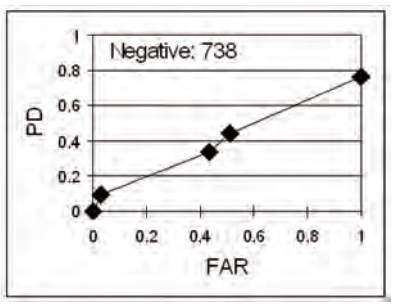


Figure 15c: ROC curve for lane 4 (testee 7). The total number of negatives (fragments, clutters or noise) is shown.

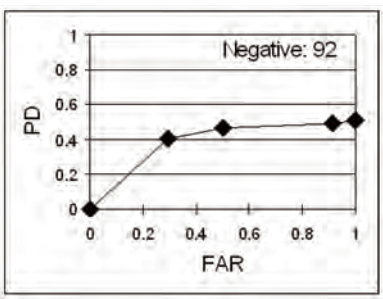


Figure 15d: ROC curve for lane 4 (testee 3). The total number of negatives (fragments, clutters or noise) is shown.

Jun Ishikawa received bachelor's and master's degrees in control engineering from the Tokyo Institute of Technology in 1989 and 1991 respectively. He joined NEC Corporation in 1991 as a Research Engineer of mechatronics and computer peripherals. He has been a Research Manager of the R&D Office for Supporting Anti-personnel Mine Detection and Removal Activities for the Japan Science and Technology Agency since 2002.

Mitsuru Kiyota received a bachelor's of engineering from KEIO University in 1971. He joined Mitsubishi Electric Corporation in 1971. He was engaged in the development of defense equipments. He was transferred to the Japan Science and Technology Agency in 2002 and is now a Research Manager of the R&D Office for Supporting Anti-personnel Mine Detection and Removal Activities.

Katsuhisa Furuta received his Bachelor and Master of Science and doctorate degree in engineering from the Tokyo Institute of Technology in 1962, 1964 and 1967, respectively. He is currently a professor, a member of the Board of Trustees and Director of the 21st Century Center of Excellence Project at Tokyo Denki University and the supervisor of the MEXT/JST Project of Humanitarian Anti-personnel Mine Detection and Removal Activities.

Jun Ishikawa
Research Manager
R&D Office for Supporting
Anti-personnel Mine Detection and
Removal Activities
Japan Science and Technology Agency
Shibuya-Property-Tokyu Bldg.
10F, 1-32-12, Higashi
Shibuya-ku, Tokyo 150-0011 / Japan
Tel: +81 3 5778 2001
Fax: +81 3 5778 5020
E-mail: ishikawa@mine.jst.go.jp
Web site: http://www.jstgo/jp/
kisoken/jirai/EN/index-e.html

Mitsuru Kiyota
Research Manager
R&D Office for Supporting
Antipersonnel Mine Detection
and Removal Activities
Japan Science and Technology Agency

Katsuhisa Furuta
Professor
Tokyo Denki University