Evaluation of Test Results of GPR-based Anti-personnel Landmine Detection Systems Mounted on Robotic Vehicles

Jun ISHIKAWA[a*], Mitsuru KIYOTA[a] and Katsuhisa FURUTA[b]

[a] Japan Science and Technology Agency, Shibuya-Property-Tokyu Bldg., 10F, 1-32-12 Higashi, Shibuya-ku, Tokyo 150-0011, JAPAN

[b] Tokyo Denki University, Ishizaka, Hatoyama-machi, Hiki-gun, Saitama 350-0394, JAPAN

ABSTRACT

This paper discusses a data analysis and evaluation of test results for anti-personnel landmine detection systems using ground penetrating radar (GPR) mounted on robotic vehicles for humanitarian demining. Six research teams from universities and industries founded by Japan Science and Technology Agency has been developed GPR systems in combination with electromagnetic induction (EMI) sensors, or metal detectors (MDs). As well as sensor technology itself, highly-accurate sensor positioning using advanced robotics technology plays a part in providing operators with clear subsurface images. The concept of the developed systems is to make no explicit alarm and to leave decision-making using subsurface images to the operators like medical doctors can find cancer by reading CT images. To evaluate these kinds of systems, a series of tests have been conducted in Japan from 8 February to 11 March 2005. Since operators' pre-knowledge of the locations of buried targets significantly influences the detection result in the tests, all the 6 lanes are designed to be suitable for blind tests using more than 200 landmine surrogates. The test results showed that combining GPR with MD can improve probability of detection (PD) around a depth of 20cm, where it is difficult to detect targets by using only a metal detector and that there is a room for further improvement in the PD by feeding back the test results to testees to learn typical target images, where targets were not able to be detected in the blind tests. It has also been learned that positioning control must be improved in scanning the ground with a sensor head, which is a key to making the best of use of MDs mounted on vehicles.

Keywords: Experimental design, test and evaluation, landmine detection, ground penetrating radar (GPR).

1. INTRODUCTION

This paper shows results of a test and evaluation for anti-personnel landmine detection systems using ground penetrating radar (GPR) mounted on robotic vehicles. Six research teams from universities and industries joined the test and evaluation managed by the Japan Science and Technology Agency (JST) [1]. There have been 4 sensor systems [2]-[5] using both GPR and metal detector (MD) evaluated here, which have been developed since October 2002 under the program of "Research and Development of Sensing Technology, Access and Control Technology to Support Humanitarian Demining of Anti-personnel Mines." This project was founded by the JST, the competent authority of which is the Ministry of Education, Culture, Sports, Science and Technology (MEXT). Developed sensor systems were evaluated being mounted 3 robotic vehicles [6]-[8] as shown in Figs. 1 to 3.

Concept of the developed systems is to make no explicit alarm and to provide operators with clear subsurface images. This means that the decision-making whether or not a shadow in the image is a real landmine is entirely left to the operator like medical doctors can find cancer by reading CT images. To evaluate these kinds of systems, a series of tests have been conducted from 8 February to 11 March 2005 in Sakaide City, Japan. Six test lanes were constructed using more than 200 landmine surrogates. Since operators' pre-knowledge of the locations of buried targets significantly influences the detection results of such systems developed here, all the 6 lanes are designed to be suitable for blind tests.

Fig. 1. Mine Hunter Vehicle.

Fig. 2. Advanced Mine Sweeper.

Fig. 3. Gryphon.

2. EXPERIMENTAL DESIGN

2.1 Test lanes and Landmine surrogates

In constructing test lanes, all the original soil was removed with a width of 2m until a depth of 0.5m in the vertical section, and the lanes were filled with controlled non-mineralized soil. The width of test lanes actually used is 1m, and mine surrogates were buried shallower than or equal to a depth of 0.3m. Feature of each lane is summarized as follows:

– lanes #1, #2, and #3 are 15m long with flat surface,

* ishikawa@mine.jst.go.jp; phone +81 3 5778-2001; fax +81 3 5778-5020

– lane #4 is 20m long, the surface of which has 15 bumps with a height of 10cm and a diameter of 60cm, and small stones are mixed into the controlled soil (Fig. 4),

– lane #5 simulates minefields in post-clearance inspection after mechanical demining, the soil of which is stirred and not packed, and

– lane #6 is wet where 10 liters of water per $1m^2$ is sprinkled 1 hour before the test starts.

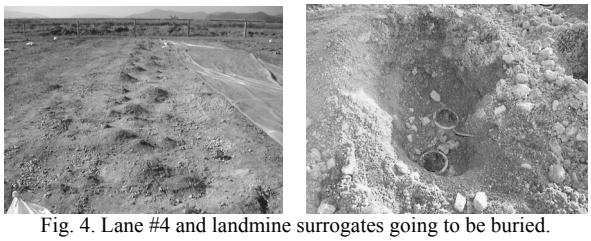
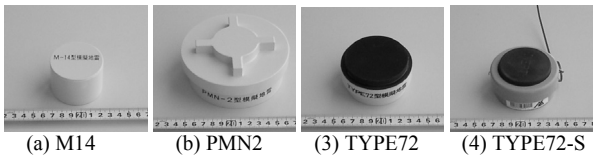


Fig. 4. Lane #4 and landmine surrogates going to be buried.

Figure 5 shows 4 kinds of landmine surrogates used in the test. M14 and PMN2 contain a metal part that is a 18mm vertical carbon steel pin with a diameter of 3mm, and TYPE72 has a 4mm vertical carbon steel pin with a diameter of 4mm. TYPE72-S is made by modifying a product of AMTECH AERONAUTICAL Ltd [9] with the exactly same part as the ITOP standard I_0, a 12.7mm vertical aluminum tube (Fig. 6). All the surrogates substitute silicone rubber as explosive.



(a) M14 (b) PMN2 (3) TYPE72 (4) TYPE72-S

Fig. 5. Landmine surrogates used in test.

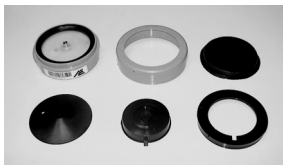


Fig. 6. Disassembled TYPE72-S surrogate.

2.2 Experimental design

Through the tests, influences of various factors on probability of detection (PD) have been evaluated. Namely, the influence of target types, target depth, soil conditions, target angle, and distance to adjacent target were considered, and 2 to 4 levels for each factor have been selected as Experiments I and II respectively summarized in Tables 1 and 2. Regarding soil condition, landmine surrogates classified into "Flat," "Wet," "Stirred," and "Rough" were respectively buried in the lanes #1 to #3, #6, #4, and #5 in a depth and angle (Fig. 7). There are 2 levels in the factor "Distance to adjacent target." One is a pair of surrogates in a distance of 15cm and the other is independent one with no other surrogates within 50cm.

Due to the limitation of time for the tests and the number of surrogates, it is impossible to test all the combinations of levels. To impartially collect data for statistical analysis under this limitation, an orthogonal experimental design based on L_{16} (2^{15}) and L_8 (2^7) orthogonal arrays was used respectively for Experiment I and II. Assigning the columns to each factor as specified in the tables derives a reduced set of combinations of levels, the results of which are summarized in Tables 3 and 4. For example, in Experiment I the number of experimental runs can reduce from 128 to 16.

According to Tables 3 and 4, landmine surrogates were buried at random locations in the lanes and had been left for more than one month before the test began. Testees can submit all the impartial data needed for analysis by reporting detection results of lanes #1 through #6.

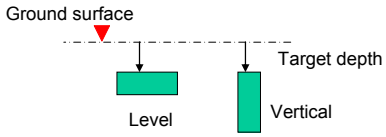


Fig. 7. Definitions of target depth and angle.

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

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 2. Factors A to C and the levels for Experiment II.

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

Table 3. Design result for Experiment I.

Experi-mental 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 4. Design result for Experiment II.

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

3. 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).
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 that is measured by TDR[1].
4. The testee reports the following data for every detected anomaly:
   - coordinates of the detected target,
   - depth of the detected target, and
   - 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 locations can be considered to be from the intended targets, that is, within a detection halo, the radius of which is half of the target diameter plus 10cm [10]. So far we do not use the reported detection depth in the judgment, but considering how to use the information is a future work.
6. Finally, the tester classifies the reported data into 4 categories as described below.
   - True positive: the case that the testee declared it as a target and this is true.
   - False positive: the case that the testee declared it as a target and this is not true. This is a false alarm.
   - True negative: the case that the testee declared it as a fragment, clutter or noise and this is true.
   - False negative: the case that the testee declared it as a fragment, clutter or noise and this is not true. This is missing a target.

Completing the tests from the lane #1 though #6 means that the testee finishes all the 16+8 experimental runs of Experiments I and II described in Tables 3 and 4.

The tester, who knows the exact locations of targets, also checked a MD response just above every buried surrogate. These data are used as a bench mark.

Table 5. Definition of confidence rating.

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.

4. DATA ANALYSIS

4.1 Analysis of variance (ANOVA)

ANOVA is to test if there are significant differences of probability of detection (PD) between levels for each factor [11]. This is for confirming that each factor is well designed to see the influence on PD. Because an objective of the test is to make limitations of the sensor systems clear, some levels such as a target depth of 30cm have been set to be very difficult in comparison with the sensor specifications.

In the following part of this section, an example is given for an ANOVA of Experiment II assuming that an experimental result in Table 6 is acquired. First the mean of results is calculated as

$$\mu = \frac{1}{8}\cdot\sum_{i=1}^{8} p_i \ , \tag{1}$$

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

$$\begin{aligned} a_{0cm} &= \frac{1}{2}\cdot(p_1+p_5)-\mu, \\ a_{10cm} &= \frac{1}{2}\cdot(p_2+p_6)-\mu, \\ a_{20cm} &= \frac{1}{2}\cdot(p_3+p_7)-\mu, \\ a_{30cm} &= \frac{1}{2}\cdot(p_4+p_8)-\mu, \end{aligned} \tag{2}$$

$$\begin{aligned} b_{Flat} &= \frac{1}{4}\cdot(p_1+p_4+p_6+p_7)-\mu, \\ b_{Routh} &= \frac{1}{4}\cdot(p_2+p_3+p_5+p_8)-\mu, \end{aligned} \tag{3}$$

$$\begin{aligned} c_{15cm} &= \frac{1}{4}\cdot(p_1+p_2+p_3+p_4)-\mu, \\ c_{MT50cm} &= \frac{1}{4}\cdot(p_5+p_6+p_7+p_8)-\mu. \end{aligned} \tag{4}$$

Next error effects e_i for $i=1,\cdots,8$ are calculated as

$$\begin{aligned} e_1 &= p_1-(\mu+a_{0cm}+b_{Flat}+c_{15cm}), \\ e_2 &= p_2-(\mu+a_{10cm}+b_{Rough}+c_{15cm}), \\ e_3 &= p_3-(\mu+a_{20cm}+b_{Rough}+c_{15cm}), \\ e_4 &= p_4-(\mu+a_{30cm}+b_{Flat}+c_{15cm}), \\ e_5 &= p_5-(\mu+a_{0cm}+b_{Rough}+c_{MT50cm}), \\ e_6 &= p_6-(\mu+a_{10cm}+b_{Flat}+c_{MT50cm}), \\ e_7 &= p_7-(\mu+a_{20cm}+b_{Flat}+c_{MT50cm}), \\ e_8 &= p_8-(\mu+a_{30cm}+b_{Rough}+c_{MT50cm}). \end{aligned} \tag{5}$$

Now, for example, a linear model for the probability of detection p_1 can be defined as

$$p_1 = \mu + a_{0cm} + b_{Flat} + c_{15cm} + e_1 . \tag{6}$$

For the ANOVA, 4 sums of squares must be calculated as follows:

$$S_A = 2\cdot\left[{a_{0cm}}^2 + {a_{10cm}}^2 + {a_{20cm}}^2 + {a_{30cm}}^2\right], \tag{7}$$

$$S_B = 4\cdot\left[{b_{Flat}}^2 + {b_{Rough}}^2\right], \tag{8}$$

$$S_C = 4\cdot\left[{c_{15cm}}^2 + {c_{MT50cm}}^2\right], \tag{9}$$

[1] The measurements of water content ranged from 6% to 16% for the lanes #1 to #5 and 9% to 22% for the lane #6.

$$S_e = \sum_{i=1}^{8} e_i^{\,2} \,. \qquad (10)$$

By comparing the variances due to levels of each factor, i.e., S_A , S_B and S_C with the variance due to measurement error S_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 are all equal. The computed F statistic in Table 7 follows an F distribution with corresponding degrees of freedom. 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.

In the experiment, to see the significant differences between testees (or sensor systems), the results was analyzed by an ANOVA of randomized complete block design (RCBD). Due to limitations of space, the details are omitted, but the experimental runs were divided into *r* equal blocks as a factor, X, where *r* is the number of testees/systems as replications.

Table 6. Notation of detection of probability.

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

Table 7. Analysis of variance (ANOVA).

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

4.2 Receiver operating characteristic (ROC) curve

It has been 30 years since radiographic applications of ROC curves reported [12] 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 (FAR) in the same way that existing landmine detectors have been evaluated.

As described in Section 3, detection results reported by testees are classified into 4 categories, i.e., true positive, false positive, true negative and false negative. However, the classification done with a testee's discrimination threshold is a one-sided view, and the number of true positive and the number of false positive are to be changing as the threshold is varied. ROC curve shows us the relationship between the true positive and false positive for a variety of different threshold, thus helping the determination of an optimal threshold as well as the comparison of sensor performance.

To plot a ROC curve, two histograms, which are measured on an interval scale in the confidence rating reported by the testee, are derived. One is from signals of intended targets, i.e., true positive and false negative, and the other is from signals of fragments, clutters or noise, i.e., true negative and false positive. After that, the ratio of true positive, i.e., probability of detection is plotted as a function of the ratio of false positive at every confidence rating (threshold). As shown in Fig. 8, 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 towards upper left hand corner of the graph as shown by solid line in Fig. 9. On the other hand, if another sensor gives a histogram of targets (dashed line), which is distributed closer to that of noise, the resulting ROC curve gets closer to a diagonal line as shown by dashed line in Fig. 9. This means that the discriminating power decreases. Once ROC curves are obtained, there are many methods to test the difference between ROC curves [13].

In the experiment, the number of true positive 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.

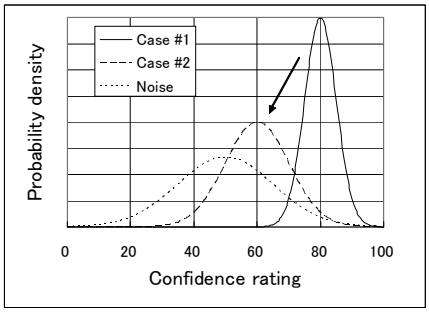


Fig. 8. Normalized histogram of signal and noise.

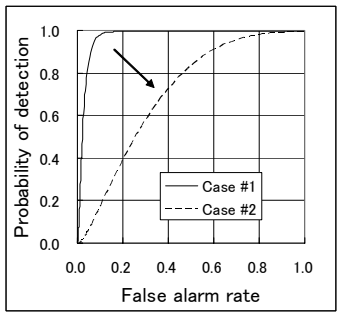


Fig. 9. Example of ROC curves.

5. EXPERIMENTAL RESULTS

Figure 10 shows the ground truth of the lane #2, and Figs. 11 and 12 shows subsurface images from a sensor system. In this case, it has been shown that a metal detector can clearly image 7 pair of Type72 surrogates buried flush (Fig. 11), and that a ground-penetrating radar can display 7 PMN2 surrogates at a depth of 20cm (Fig. 12), 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.

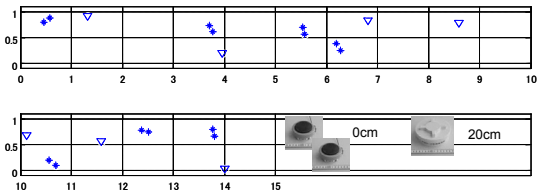


Fig. 10. Ground truth of the lane #2: ** shows a pair of Type72 and ∇ shows PMN2.

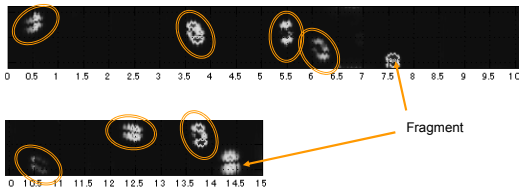


Fig. 11. Detection image from a metal detector.

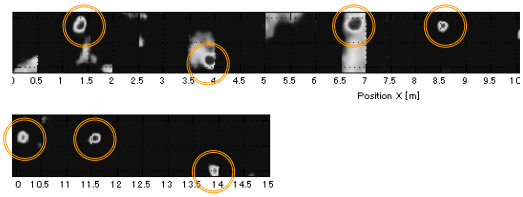


Fig. 12. Detection image from a GPR.

5.1 Probability of detection

The number of testees is ten, the breakdown of which are 3 of MHV with a pulse GPR+MD, 2 of MHV with a step-frequency GPR+MD, 3 of AMS with a step-frequency GPR+MD, and 2 of Gryphon with a pulse GPR+MD. One testee from every system has been chosen, and the 4 sets of data were analyzed by ANOVA, where 4 testees/systems was treated as a factor, X, of ANOVA of RCBD.

Tables 8 and 9 respectively show ANOVA results for Experiments I and II. 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 system. Regarding the factor, C, of Experiment II, i.e., distance to adjacent target, the ANOVA showed that there was no significant difference in PD between a pair of TYPE72-S surrogates in 15cm distance and the other independent ones.

Average PD of 10 testees for each level of the factor is depicted in Figs. 13 and 14 in comparison with the MD bench mark explained in Section 3. In these figures, highest and lowest PD are also plotted. These figures showed that there is a strong dependence of target depth on PD as well as the result using only a metal detector, but PD in deep levels of 20-30cm can be improved by combining GPR with MD. On the other hand, as also shown in Figs. 13 and 14, GPR+MD results in shallow levels of 0-10cm were worse than those of MD. This is because sensor height to the ground that controlled by robot arms was higher than that of manual scanning of MD. It is known that GPR has a difficulty in imaging shallow targets and that MD, which is very sensitive to standoff distance, plays an important role in detecting shallow targets. Therefore, the sensor height control is a key to solving the problem. For the same reason, PD in rough surface was not improved from MD results, and this will be also solved by fixing a robotic part, which will be a future work.

Table 8. Result of ANOVA of Experiment I.

Source of Variation	Degree of freedom	Sum of squares	Mean of squares	Computed F statistic	Probability at the F value
A: Target type	3	0.3989	0.1330	3.1053	0.0347 *
B: Target depth	3	3.4115	1.1372	26.5580	0.0000 **
C: Soil condition	3	1.1973	0.3991	9.3211	0.0001 **
D: Target angle	1	0.4850	0.4850	11.3273	0.0015 **
X: System (Testee)	3	0.6667	0.2222	5.1904	0.0034 **
e: error	50	2.1409	0.0428		
Total	63	8.3003			

Table 9. Result of ANOVA of Experiment II.

Source of Variation	Degree of freedom	Sum of squares	Mean of squares	Computed F statistic	Probability at the F value
A: Target depth	3	1.1000	0.3667	9.2485	0.0003 **
B: Soil condition	1	0.6328	0.6328	15.9617	0.0006 **
C: Distance to adjacent target	1	0.0270	0.0270	0.6800	0.418
X: System (Testee)	3	0.4570	0.1523	3.8425	0.0229 *
e: error	23	0.9118	0.0396		
Total	31	3.1286			

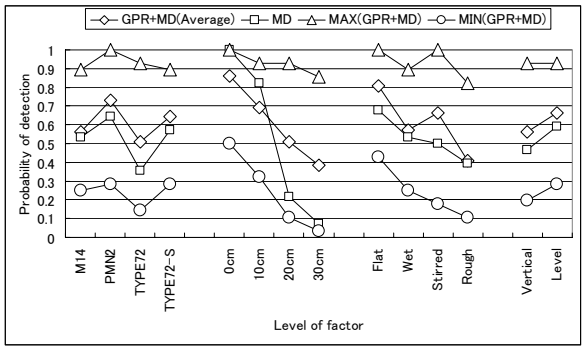


Fig. 13. Average PD for each level of factor of Experiment I.

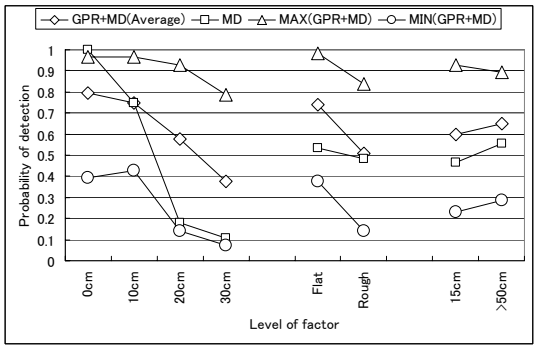


Fig. 14. Average PD for each level of factor of Experiment II.

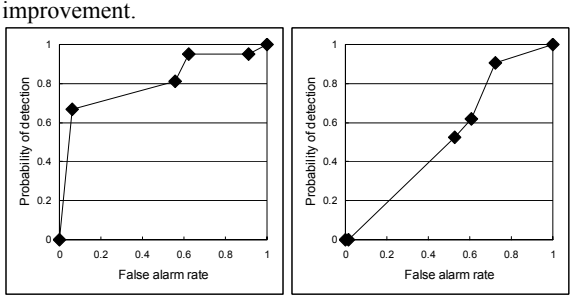
5.2 False alarm evaluation

High probability of detection (PD) has to be followed by low false alarm rate (FAR). In a system such that an operator

makes a decision to detect targets, confidence rating plays an important role in evaluation of the system effectiveness seeing the relationship between PD and FAR. From a view point of this, well-functioning as a sensor can be defined by a qualification such that high PD is attained and

- many objects, i.e., fragments, clutters, noise, are detected as well as intended targets, but two histograms measured in the confidence rating are clearly separated, or
- the number of false alarms is very few no matter how the histograms are distributed.

As described in Section 4.2, ROC curves are useful to see the qualification. Figures 15(a) to 15(d) show typical ROC curves for the lane # 2 where 21 targets were buried as shown in Fig. 10. A horizontal axis of each plot shows false alarm rate (FAR), and the number of false alarm is calculated by FAR multiplied by the total number of negatives that is indicated in each figure caption. In the case of Fig. 15(a), 65% of targets were quite obviously detected, but the others got mixed in 525 negatives. Figure 15(b) showed that the testee could not discriminate the target from 520 negatives. In Fig. 15(c), almost all the target was detected at 100% confidence, and Fig. 15(d) showed that 85% of targets were detected with a very few false alarms. These results were fed back to the testees for further improvement.



(a) Positive: 21 /Negative: 525 (b) Positive: 21/ Negative: 520

(c) Positive: 21/ Negative: 71 (d) Positive: 21/ Negative: 17

Fig. 15. Some typical ROC curves of the lane #2: caption of each figure shows the total number of positives (targets) and the total number of negatives (fragments, clutters, or noise)

6. CONCLUSIONS

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

- PD in deep levels of 20-30cm can be improved by combining GPR with MD.
- Provided calibration area should have contained landmine surrogates for all the levels of factors. Coaching a typical image for each level would much improve the detection rate.
- In some case, high PD have been accompanied by many false alarms, and it was also proven that confirming the source of GPR false alarms is much more difficult than those of MD. Therefore, another performance index that penalizes GPR false positives will be needed in GPR evaluation.

The most important thing is to use these technologies to improve mine detection efficiency and reduce minefields, and the mine detection systems must be robust, simple and highly cost-effective. After reflecting the lessons and improving the sensor systems, the next step of the project is field tests to evaluate these features in some mine-affected countries.

REFERENCES

[1] J. Ishikawa, M. Kiyota, and K. Furuta, "Experimental design for test and evaluation of anti-personnel landmine detection based on vehicle-mounted GPR systems," *Proc. of SPIE Vol. 5794, Detection and Remediation Technologies for Mines and Minelike Targets X*, pp. 929-940, 2005.

[2] M. Sato, Y. Hamada, X. Feng, F. Kong, Z. Zeng, G. Fang, "GPR using an array antenna for landmine detection," *Near Surface Geophysics*, Vol. 2, pp. 3-9, February, 2004.

[3] M. Sato, J. Fujiwara, X. Feng, Z. Zhou, and T. Kobayashi, "Development of a hand-held GPR MD sensor system (ALIS)," *Proc. of SPIE Vol. 5794, Detection and Remediation Technologies for Mines and Minelike Targets X, pp. 1000-1007, 2005.*

[4] S. M. Shrestha and I. Arai, "High Resolution Image Reconstruction by GPR using MUSIC and SAR Processing Method for Landmine Detection," *Proc. of the 2003 IEEE International Geoscience and Remote Sensing Symposium* (*IGRASS2003*), 2003.

[5] Y. Hasegawa, K. Yokoe, Y. Kawai, and T. Fukuda, "GPR-based Adaptive Sensing – GPR Manipulation According to Terrain Configurations –," *Proc. of the 200 IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS2004*), 2004.

[6] K. Nonami, and H. Aoyama, "Research and Development of Mine Hunter Vehicle for Humanitarian Demining," *Proc. of the IARP International workshop on Robotics and Mechanical Assistance in Humanitarian Demining* (*HUDEM2005*), to appear.

[7] T. Fukuda, et al., "Environment-Adaptive Antipersonnel Mine Detection System - Advanced Mine Sweeper -," *Proc. of the HUDEM2005*, to appear.

[8] E. F. Fukushima, et al., "Teleoperated Buggy Vehicle and Weight Balanced Arm for Mechanization of Mine Detection and Clearance Tasks," *Proc. of the HUDEM2005*, to appear.

[9] K.R. Torrance, C.G. Coffey, A.B. Markov and W.S.Myles, "Surrogate AP Mines for Training Deminers and Evaluating Demining Equipment," *Proc. of the EUDEM2-SCOT2003 (International Conference on Requirements and Technologies for the Detection, Removal and Neutralization of Landmines and UXO)*, 2003.

[10] CEN Workshop Agreement, *Humanitarian Mine Action - Test and Evaluation- Metal Detectors*, CWA 14747, 2003.

[11] K. Jayaraman, A Statistical Manual for Forestry Research, Food and Agriculture Organization of the United Nations Regional Office for Asia and the Pacific, March 1999 (http://www.fao.org/documents/show_cdr.asp?url_file=/DOCREP/003/X6831E/X6831E00.HTM).

[12] D. J. Goodenough, K. Rossmann, and L. B.Lusted, "Radiographic applications of receiver operating characteristics (ROC) curves," Radiology, vol. 110, pp. 89-95,1974.

[13] C. E. Metz, B. A. Herman, and C. A. Roe, "Statistical Comparison of Two ROC-curve Estimates Obtained from Partially-paired Datasets," *Medical Decision Making*, 18, pp. 110-121, 1998.