Comparison of ROC curves for landmine detection by holographic radar with ROC data from other methods

T. Bechtel
Dept. of Earth and Environment
Franklin & Marshall College
Lancaster, PA, USA
tbechtel@fandm.edu

C. Windsor
116 New Road
East Hagbourne
Oxfordshire, UK

L. Capineri
Dept. of Information Engineering
University of Florence
Florence, Italy

M. Inagaki
Walnut Ltd.
Saiwaicho, Tachikawa
Tokyo, Japan

S. Ivashov
Remote Sensing Laboratory
Bauman Moscow State Technical University
Moscow, Russia

Abstract* — In de-mining or UXO work, classification of subsurface targets into bins such as "mine" vs. "clutter" is critical. So is statistical evaluation of classification accuracy. For radar systems with automated target recognition involving some threshold parameter, one can plot an ROC curve showing the detection rate versus the false alarm rate across a range of threshold values. However, for visual interpretation of images, there is no parameter. Instead, the operator makes a "judgment call" based on training and experience. We propose that for visual interpretation of radargrams, differences in judgment between operators are a proxy for a variable threshold parameter. To test this, we recorded holographic radar images of a test bed containing plastic mine casings and clutter items. Each image contained between 0 and 3 mines, and 3-7 clutter items. University students with no prior training in mine or UXO recognition where given minimal training, and then asked to interpret the images. The detection and false alarm rate for each test subject across all of the images yielded a single point on an ROC curve. In addition, the false alarm rate for each clutter item was determined individually. Based on this, it appeared that rounded rocks are the most frequent false alarm. ROC curves for the "worst operators" were compared to published ROC data from other landmine detection methods, and fall within the range of performance for these other methods – even for testing by trained operators. We propose that in appropriate conditions, the holographic method will provide competitive detection metrics, even by minimally-trained lay-people such as de-miners recruited from within mined communities, and that the described method for developing ROC data can be used to quantify their performance with statistical significance.

Keywords—holographic radar, de-mining, UXO, probability of detection, probability of false alarm, ROC curve

I. INTRODUCTION

RASCAN is a holographic radar designed by the Remote Sensing Laboratory, Bauman Moscow Technical University [1] originally for use in humanitarian demining [2], but also adapted for use in construction [3], architectural or cultural preservation [4], aeronautics [5], and paleontology [6]. For the original mine detection purpose, laboratory trials have provided proof-of-concept [7], shown that the radar images are suitable for automated target recognition [8], and can be recorded using a low-cost, lightweight robotic scanner [9]. However, the performance of RASCAN for landmine detection has not been evaluated statistically. This paper reports on mock field trials designed to develop robust probability of detection (PD) or true positive rate for several difficult mines in the presence of significant clutter, and to generate receiver operating characteristic (ROC) curves [9] relating PD to probability of false alarm (PFA) or false positive rate.

Since our work has focused on developing a simple, low-cost, highly portable landmine detector, we tested the capabilities of a RASCAN-4/4000 real-time subsurface imaging system with no associated data or image post-processing. The RASCAN-4/4000 produces 10 simultaneous plan-view images (5 discrete frequencies from 3.6 to 4.0 GHz, using receivers both parallel and perpendicular to the transmitter) that appear in real-time on a Windows™ laptop or tablet screen as the sensing head is swept in a raster fashion across the tested surface. Upon completion, the software combines these into an animated image that morphs the 10 images in a continuously looping display which the operator interprets visually. The purpose of this test was to determine how accurately minimally-trained operators could identify mines whilst rejecting clutter.

II. METHOD

The RASCAN device has no sensitivity setting, and for this study, the images were interpreted visually with no algorithm using adjustable parameters. Therefore, the variable judgment of a large number of operators was employed as a proxy for sensitivity. The subjects were 70 college students representing a variety of fields of study, but none with any prior landmine or UXO or GPR experience or knowledge. We expected that within this group there should be some variability in "sensitivity" but to ensure a wide range, prior to testing, 18 students were secretly given one of two sets of instructions on a slip of paper:

1-"You are a high sensitivity detector. We cannot afford to miss any mines, so be aggressive in your identifications. If you even suspect a mine, label it as a mine."

2-"You are a low sensitivity detector. If we spend our entire budget digging up clutter, we will have nothing left to remove actual mines. Do not identify a target as a mine unless you are absolutely certain it is a mine."

Thus, approximately 13% of subjects were high sensitivity, and 13% were low sensitivity, while the rest were simply instructed to use their own judgment.

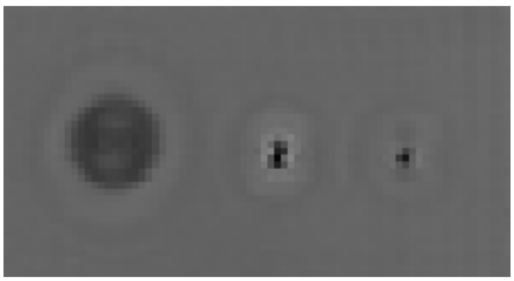

Figure 1: RASCAN subsurface holographic radar image, 3.6 GHz, parallel polarization.

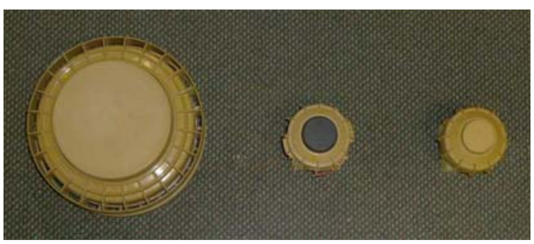

Figure 2: Three plastic body mines; VS-1.6 (Φ=242 mm), VS-50 (Φ =90 mm), TS-50 (Φ =90mm).

Training, immediately followed by testing, was done using subjects in 3 groups (25, 25, and 20 students). Training involved collective viewing of a single animated RASCAN image of 3 mines (Fig. 1), as well as a matching photograph of the arrangement of the mines in the test bed (Fig. 2). Students also handled inert versions of the mines in order to instill a better intuitive sense of their shapes and dimensions. Total training for each group lasted less than 30 minutes.

Following training, students worked entirely on their own to identify mines on 8 testing images. The scans were pre-recorded in an outdoor test bed filled with long-graded builders' sand at natural moisture content (~8 percent by weight). Each of the 8 testing scans had 0, 1, 2, or 3 mines, plus between 5 and 10 clutter items, with a minimum clutter:mine ratio of 2.5:1. Fig. 3 presents still frames from two example scans, along with their matching object layout photographs (not shown to the subjects). Clutter objects were chosen to represent typical war zone and/or urban detritus. The mines all had plastic casings, minimum metal content, and filling having a dielectric constant close to plastic explosives, making them difficult targets for detection. All objects were buried at depths between 2 and 10 cm.

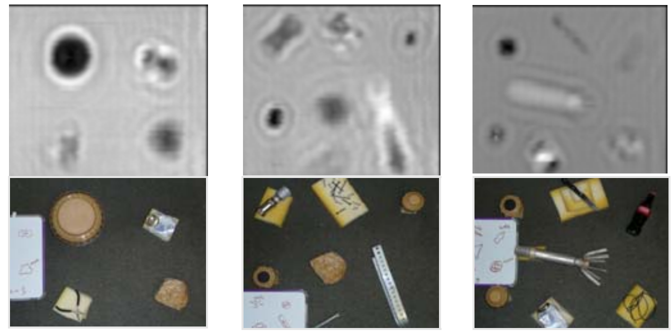

Figure 3: Top - example RASCAN images (75x65 cm); Bottom - matching object layouts as buried in test bed.

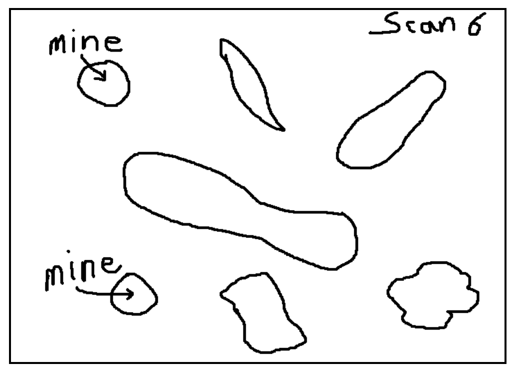

Figure 4: One subject's sketch based on the RASCAN image in the upper right of Fig. 3.

Setup of the test bed, and scanning of the training and test images was performed by a separate group of students prior to analysis by the test subjects, with no communication between these groups to ensure a blind test. Scanning of each 0.75m x 0.65m image took less than 3 minutes.

For the image analysis phase, each student was shown each testing image for 2 minutes, and instructed to draw a sketch of their interpretation, labelling items that they identified as mines. An example sketch, for the RASCAN image and layout on the right hand side of Fig. 3, is shown in Fig. 4. In total, testing involved 510 image analyses representing 249 m^2 of "lane", containing 3060 clutter items interspersed with 810 mines. The interpretive sketches were collected at the end of viewing to ensure no post-viewing "second-guessing".

III. Data Analysis

Investigators scored the sketches by comparison with the layout photos. An object was considered detected if there was a shape with a roughly correct outline and orientation, in the correct position relative to all other shapes. Targets were classified as “missed” if there were fewer shapes on the sketch than objects in the photo. In all but a few instances, it was easy to tell which target had been missed, and the few cases where it was ambiguous are not considered important since there were so few total misses. Fig. 5 shows the detection rate for all clutter items, along with the rates at which they were identified as mines (false alarms). Note that the overall detection rate for clutter items was above 98%, and that the 3 rocks accounted for ~50% of false alarms. Presumably this is because the rocks are equant and non-metallic so they mimic the plastic mines in the radar images. The crushed can also produced significant false alarms – again perhaps due to its equant shape.

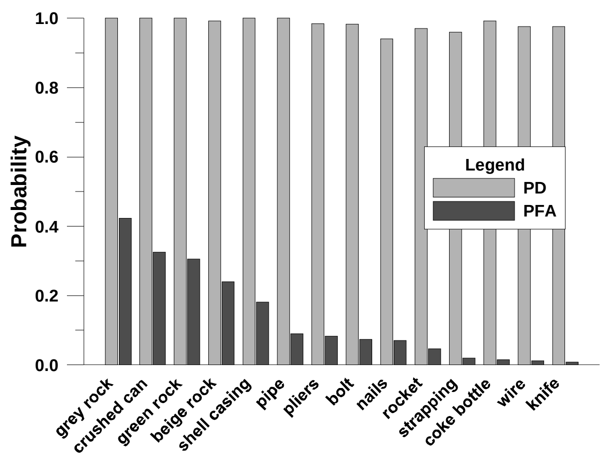

Figure 5: Detection and false alarm probabilities for clutter items.

For the mines, the raw probabilities of detection were as follows: **TS-50; 87%, VS-50; 90%, VS-1.6; 99%**. However, following the derivation of Simonson [11], these need to be corrected for sample size, and can only truly be presented as upper and lower limits based on some specified confidence level. At a 99.5% confidence level (P=0.05), the appropriate values from this test are: **TS-50; 83% to 90%, VS-50; 87% to 93%, VS-1.6; 97% to 99%.**

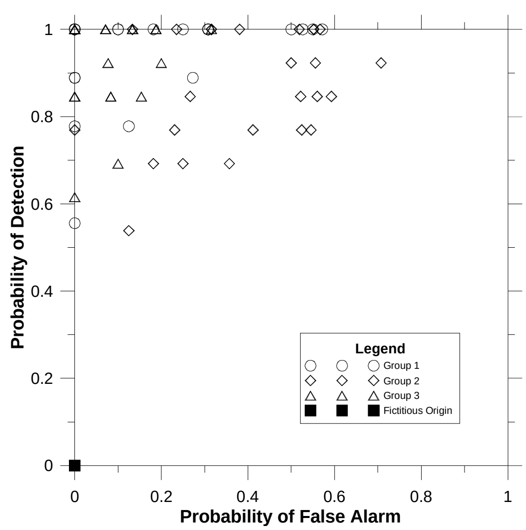

Figure 6: ROC data for all groups, all subjects and all images.

A more complete characterization of the combined detector and operator performance is the ROC curve data depicted in Fig. 6 where the PFA versus PD data are plotted for all subjects. Notice that the data generally follow a typical ROC pattern, with PFA increasing more rapidly as PD increases. Note also that there is a wide range of PFA values indicating the expected range of judgment or aggressiveness in mine identification that was probably accentuated by the specific instructions to certain subjects.

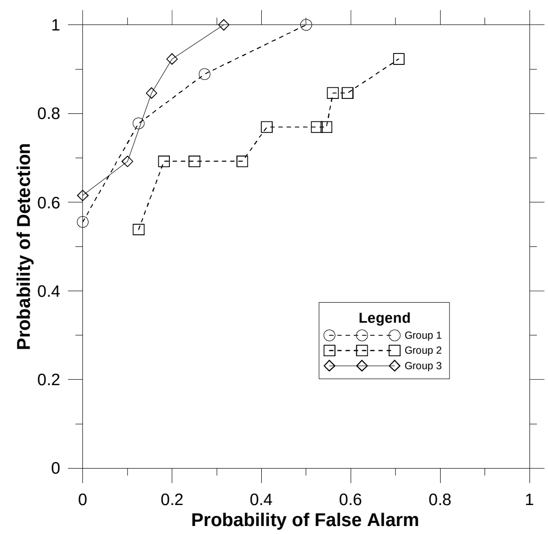

Figure 7: ROC curves for "worst operators"

The data from the three groups of subjects are plotted with differing symbols because this makes apparent a systematic difference between them. This is even more obvious on Fig. 7, where only the worst performers from each group are used to make a single ROC curve. Worst performers were defined as the lowest PD for any given PFA. While Groups 1 and 3 achieved 100% PD at 51% and 32% respectively, with overlap between their respective curves, the worst performers in Group 2 achieved a best PD of only 92%, and achieved consistently lower PD at all PFAs. With the sets of images being identical, the pool from which subject groups were drawn being the same, and the room used for testing remaining unchanged, the difference may be related to transient human factors – the effectiveness of the seemingly-identical training or level of subject distraction related to weather, hunger, day of the week, etc. The importance of these factors is already well recognized [e.g. 12], but particularly highlighted by these data.

If the systematically poorer Group 2 data are eliminated, the detection statistics change from those reported above to the following: **TS-50; 89% to 97%, VS-50; 83% to 93%, VS-1.6; 94% to 100%**. Note that the overall statistics do not improve significantly. The main effect is to broaden the range between the upper and lower estimates due to a smaller sample size. Thus, while the ROC appears to be worse for Group 2, this difference may not be statistically significant.

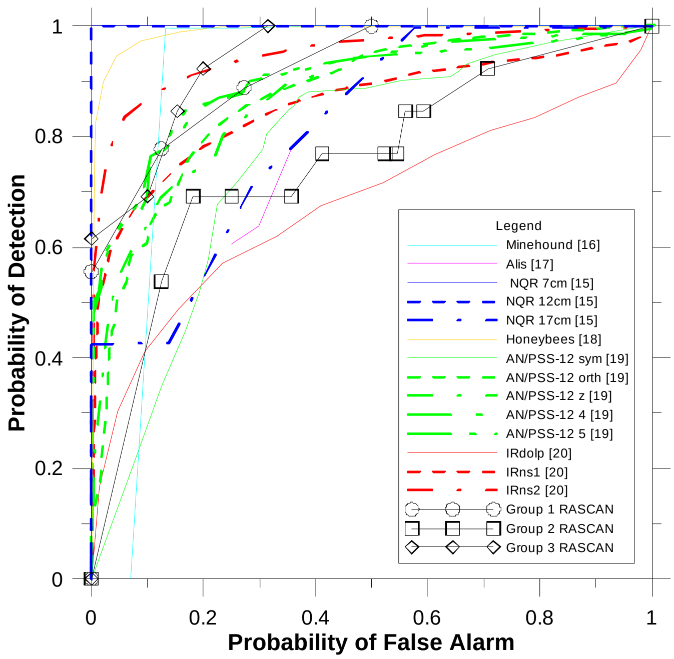

Figure 8: Comparison of RASCAN "worst operator" ROC curves with other published data.

Since detector performance is strongly tied to myriad parameters including human factors, environmental factors (e.g. soil type and moisture), and target factors (e.g. clutter density and mine types), comparisons between different detectors in different field tests are not generally valid [13]. However, in order to place the results of these holographic radar tests in some context, Fig. 8 overlays the worst operator results cited above on ROC data from the literature for a range of detector types. This crude comparison provides motivation to continue this work. Groups 1 and 3 with their minimal training and total inexperience performed quite favorably compared to other methods wielded by much more expert users. Even the underperforming Group 2 manages to fall within the range of some of the other systems. We are confident that with additional experience, and in combination with the metal detector coil as intended in the original MiRASCAN system [1], operators could perform as well as any of the systems represented here.

It is also important to note two things. First, because of the range of factors influencing detector performance, no single tool can be effective in all demining settings, thus it will always be "necessary to apply a process with complementary tools and techniques." [J. MacDonald, 13]. Second, "the stereotype image of a flat, grassy minefield is misleading and harmful… Yet the image is constantly reinforced by the trials, demonstrations, and publicity shots that invariably take place in near-perfect conditions" [C. King, 14]. This means that while we have shown that untrained and inexperienced operators can quickly and confidently interpret RASCAN images collected in test beds, a significant challenge will be construction of a system that maintains these characteristics in far less hospitable terrain. Thus, our immediate future work will focus on the effects of relief, vegetation, and soil moisture on RASCAN images and ROC curves.

Acknowledgments

This work was partially supported by the GeoScience Founders Society, Franklin & Marshall College. Enviroscan, Inc., Lancaster, PA generously provided equipment and laboratory space, and the use of their outdoor test bed.

References

[1] Ivashov, S. I., L. Capineri, and T. D. Bechtel. "Holographic Subsurface Radar Technology and Applications, in book UWB Radar Applications and Design, edited by JJ Taylor, 421–444." (2012).

[2] Ivashov, Sergey I., et al. "Remote control mine-detection system with GPR and metal detector." 8th International Conference on Ground Penetrating Radar. International Society for Optics and Photonics, 2000.

[3] Ivashov, S. I., et al. "Holographic subsurface radar as a device for NDT of construction materials and structures." Nondestructive Testing of Materials and Structures. Springer Netherlands, 2013. 799-804.

[4] Capineri, L., et al. "Application of the RASCAN holographic radar to cultural heritage inspections." Archaeological Prospection 16.3 (2009): 218-230.

[5] Ivashov, Sergey I., et al. "Comparison between impulse and holographic subsurface radar for NDT of space vehicle structural materials." PIERS Online3.5 (2007): 658-661.

[6] Capineri, L., et al. "RASCAN holographic radar for detecting and characterizing dinosaur tracks." Advanced Ground Penetrating Radar (IWAGPR), 2013 7th International Workshop on. IEEE, 2013.

[7] Capineri, L., et al. "Comparison of GPR sensor types for landmine detection and classification." 12th International Conference on Ground Penetrating Radar. Vol. 55. 2008.

[8] Windsor, C., L. Capineri, and T. D. Bechtel. "Buried object classification using holographic radar." Insight-Non-Destructive Testing and Condition Monitoring54.6 (2012): 331-339.

[9] Capineri, L., et al. "High Resolution Imaging with a Holographic Radar Mounted on a Robotic Scanner." Session 4A5 SC5: High Resolution Imaging with Penetrating Radar Scanners for Detection of Small or Low Contrast Objects: 1378.

[10] Metz, Charles E. "Basic principles of ROC analysis." Seminars in nuclear medicine. Vol. 8. No. 4. WB Saunders, 1978.

[11] Simonson, Katherine M. "Statistical considerations in designing tests of mine detection systems: I-Measures related to the probability of detection." Sandia National Laboratories. 1998.

[12] Bacchi, M., et al. "Human Factors and Professional Competence in Humanitarian De-mining." wpeati Öjßce o. (2000): 40.

[13] MacDonald, Jacqueline, et al. Alternatives for landmine detection. No. RAND/MR-1608-OSTP. RAND CORP SANTA MONICA CA, 2003.

[14] King, Colin. "Demining: enhancing the process." (1998): 1-9.

[15] Itozaki, Hideo. "Nuclear quadrupole resonance for explosive detection." Anti-personnel Landmine Detection for Humanitarian Demining. Springer London, 2009. 147-155.

[16] Daniels, David J. "Ground penetrating radar for buried landmine and IED detection." Unexploded Ordnance Detection and Mitigation. Springer Netherlands, 2009. 89-111.

[17] Takahashi, Kazunori, and Motoyuki Sato. "A hand-held dual-sensor system using impulse GPR for demining." Ultra-Wideband, 2008. ICUWB 2008. IEEE International Conference on. Vol. 3. IEEE, 2008.

[18] Rodacy, Philip J., et al. "Training and deployment of honeybees to detect explosives and other agents of harm." AeroSense 2002. International Society for Optics and Photonics, 2002.

[19] Collins, Leslie, Ping Gao, and Lawrence Carin. "An improved Bayesian decision theoretic approach for land mine detection." Geoscience and Remote Sensing, IEEE Transactions on 37.2 (1999): 811-819.

[20] Gurton, Kristan P., and Melvin Felton. "Remote detection of buried land-mines and IEDs using LWIR polarimetric imaging." Optics express 20.20 (2012): 22344-22359.