Multi-sensor threat detection for screening people and their carried bags

Samuel Pollock, Daniel R. Crick, Laura J. Winter, and Michael C. Kemp

Iconal Technology Ltd., St John's Innovation Centre , Cambridge, United Kingdom

ABSTRACT

Terrorists increasingly target crowded places, such as sporting and entertainment venues, visitor attractions and transport hubs, when planning mass-casualty attacks. The organisations responsible for these venues must therefore consider how to efficiently screen visitors and their carried bags for potential threats. In particular, screening at these sites needs to be carried out quickly, at low cost and with minimal interruption to the normal flow of commerce. We describe a multi-sensor threat detection system based on hybrid electromagnetic, ultrasound, microwave and optical detection techniques, which provides new screening capabilities and automated detection at low hardware and operating costs. Its compact size also enables deployment in many different locations and application areas. This system is one of very few technologies that can screen both people and their carried bags simultaneously for the presence of both metallic and non-metallic threats.

Keywords: Explosive detection, multi-modal, sensor fusion, microwave, millimeter wave, stand-off, ultrasound, threat detection

1. INTRODUCTION

This paper reports development of an integrated multi-sensor system to detect both metallic and non-metallic mass-casualty threats concealed on a person's body or in a backpack or other carried bag. Much of the work has focused on detection of threats in carried bags – a desirable capability that is poorly addressed by existing stand-off and walk-by people screening systems.

Mass-casualty threats used in crowded places are likely to be relatively bulky, so bags containing threats are expected to be full and contain few other items, making them relatively homogeneous. In comparison, benign bags are likely to vary in fullness and contain a variety of items. This contrasts with aviation security, where threats may be small relative to the size of the bag and concealed among many benign items. The system outlined in this report, named AcES, uses these characteristic "signatures" of threat and harmless items to optimise its performance to its target application.

Our vision is that the system will primarily act as a first-line filter or triaging tool. It will operate at high throughput to clear the majority of people with bags that are very unlikely to contain a large weapon or person-borne improvised explosive device (PBIED). This provides a non-intrusive, fast and evidence-based method of selecting people for screening by other technologies or manual search in high-throughput applications. Previous attempts to address the challenge of screening undivested people and their bags have focused on trying to design the "perfect" single sensor. We recognised that no single technology can address the full detection challenge and therefore adopted a multiple sensor approach, focusing on the characteristic signatures of mass-casualty threats in crowded places. We integrated our sensors into a single unit, unlike previous multi-sensor approaches, thus reducing space requirements and cost.

We aimed to identify a suite of complementary sensors that measured a wide range of properties of bags and their contents, such as their mass, volume, dielectric constant, metal content and homogeneity. These properties would be combined to form signatures indicating whether a person and their bag were benign or could contain a concealed threat.

Email: sam.pollock@iconal.com; Web: http://www.iconal.com

2. METHODS

2.1 Sensor selection

Where possible, we focussed on adopting widely-available, low-cost, commercial off-the-shelf (COTS) components developed for other (particularly mass market) applications. For example, 3D imaging is heavily exploited in gaming and chip-scale microwave and millimetre-wave radar is widely used in the automotive industry and industrial sensing. This reduced both the potential cost and the development time of the system by eliminating costly and time-consuming bespoke sensor engineering.

A wide range of technologies and components was explored and down selected to a core set of complementary sensing modalities:

- 3D optical imaging
- microwave imaging radar
- millimetre-wave employing ranging, polarimetric and micro-Doppler sensing
- ultrasonic sensing.

These modalities are described individually in the following sections.

2.2 3D optical imaging

A 3D optical imaging camera was employed to record: the real-time position of a subject relative to the sensor, their orientation and pose, the presence, dimensions and orientation of any bag they were carrying, and an image of the subject (if necessary), which could be passed to a security staff member for identification.

The most critical function of this sensor is providing positional information to calibrate, coordinate and align responses from all the other sensors. The absolute position of the back of the subject relative to the bag is also important, as this determines whether a pulse (radar or ultrasound) has passed straight through the bag (and reflected off the torso) or has reflected off something within the bag. Figure 1 shows examples of images from the 3D camera.

The Microsoft Kinect One Time-of-Flight camera was used in early prototypes, but this was discontinued in late 2017. We therefore used an alternative 3D camera based on the structured-light principle in later phases of the work. A 3rd-party software package provides real-time pose information (skeletal tracking) for subjects in the field of view.

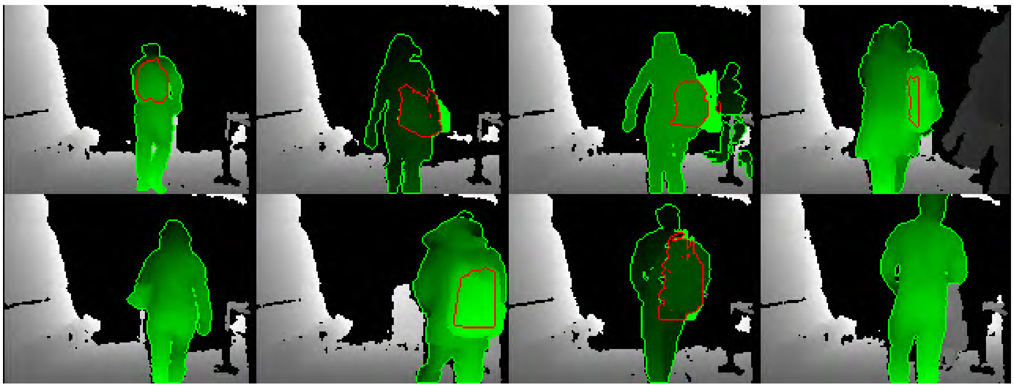

Figure 1. Series of 3D camera images of subjects (outlined in green), some carrying bags (highlighted in red when detected on the back).

2.3 Microwave and millimetre-wave radar

A microwave (3–10 GHz region) 3D imaging radar is used to create a low-resolution 3D radar map of the subject in motion. The 3D radar module is based on a commercially-available Walabot Developer device, which uses a single chip radar developed by Vayyar and an array of 18 antennas. The measurements from multiple transmit–receive antenna pairs are analysed to reconstruct a three-dimensional "image" of the environment.

This identifies the locations of radar significant objects in a carried bag or under clothing, and the degree to which they reflect radar energy relative to the torso of the subject. The radar probes the average properties of the bag, rather than resolving discrete layers within the area under inspection, as the wavelengths are on the several-centimetre scale. However, the ability to resolve in both the azimuth and elevation directions, as well as depth, enables localised inspection of specific regions of the bag – a significant difference from the millimetre-wave, point-sensor radar described below.

Figure 2 shows three situations: a benign subject, where the reflected radar energy is centred around the torso of the subject; a subject with a pipe-bomb vest, where the centre of mass of the radar return is shifted to the front of the torso, and, finally, a situation where the radar return is apparently shifted behind the torso. This counter-intuitive effect is caused by the delay in propagation introduced when microwaves propagate through a microwave-transparent material with a significant dielectric constant (in this case, plastic explosive simulant).[1]

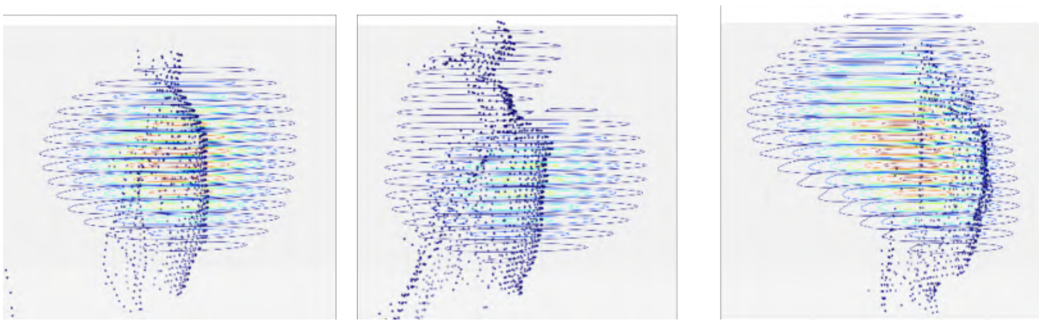

Figure 2. 3D radar images returned from subjects without a threat, wearing a pipe-bomb vest and carrying a plastic-explosive simulant in a bag(left to right). The radar propagates from the right of each image. Dotted marks indicate the position of the subject, as shown by the 3D camera. The apparent shift in position of the explosive threat is due to propagation delays through the dielectric material.

The microwave imaging radar is complemented by a millimetre-wave radar operating as a point sensor. The shorter wavelengths used in this sensor interact more strongly with the bag materials, objects within the bag, clothing layers and objects under clothing. A wide bandwidth ensures good depth resolution, which enables identification of individual layers within bags and clothing. It also enables the positions of bag or clothing surfaces relative to the torso to be identified more accurately, compared to the microwave imaging radar.

We initially used a 57–64 GHz chip-scale, frequency-modulated, continuous-wave (FMCW) radar, but we subsequently adopted a 24–26 GHz FMCW radar platform based on the availability of development kits at the time, which gave improved performance. Further development phases may return to a 60 GHz radar, due to a recent plethora of new offerings in the marketplace and to avoid the usage restrictions of frequencies in the K-band.

The 24 GHz radar module has one transmit and two coherent receive channels. It uses three external horn antennas (approx. gain 15 dBi), in which the transmit lens is coupled to a PTFE microwave lens to improve directionality, giving an inspection region approximately 200–300 mm in diameter at 2 m. One of the two receive channels is orientated in line with the transmit antenna (receiving co-polarised signal) and one is orientated at 90° (receiving cross-polarised signal) for polarimetric measurements, as outlined below.

We see similar phenomenology to that of the microwave radar (Figure 3), in that strong reflectors shift the majority of the radar return closer in range, to positions inside the bag or directly at the surface of clothing. Conversely, large quantities of semi-transparent dielectric material introduce propagation delays, shifting the

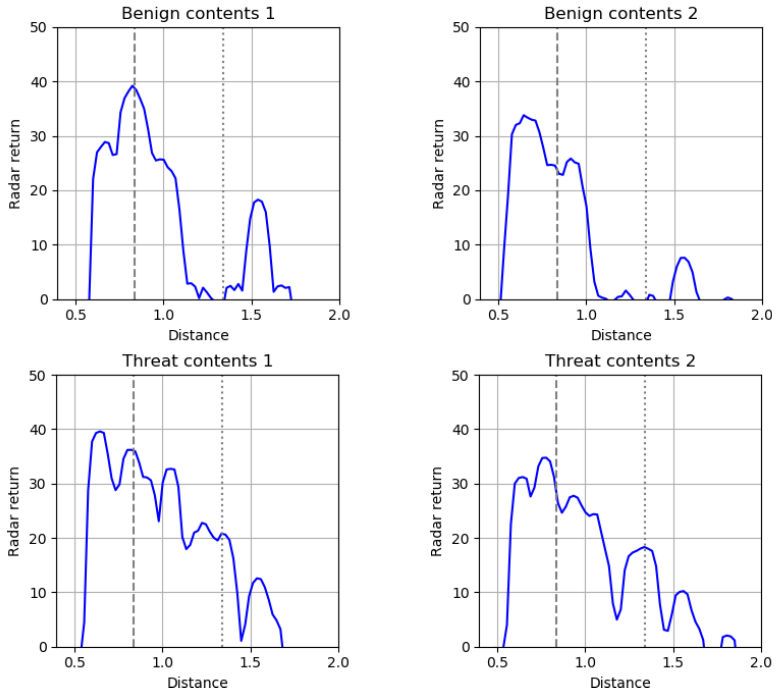

Figure 3. Millimetre-wave FMCW radar returns (using original 60 GHz radar) from two benign bags (top row) and two threat bags (bottom row). Approximate position of the back of the subject's torso is illustrated with a dashed line.

centre-of-mass of radar returns further away in range. Fusing these results with the 3D optical observations provides a direct reference to the surface of the bag or clothing, and relative positions of the torso are inferred from skeletal tracking.

The two orthogonal receive channels enable polarimetric measurements to detect metallic features.[2] "Well-behaved" reflectors, such as people and most common items, reflect radar with nearly the same polarisation (co-polar), but metal objects with sharp edges and small features causing multiple reflections can result in polarisation with a significant cross-polar component.

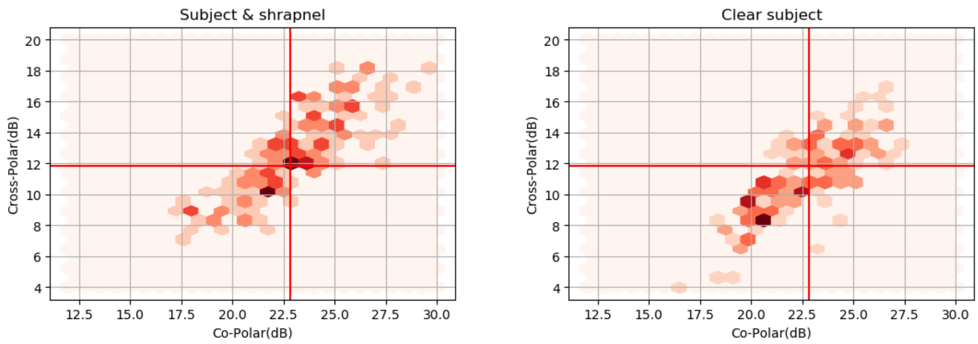

Figure 4. Distribution of reflected radar power for a subject with (left) and without (right) metallic shrapnel. Red lines indicate the average cross-section across both data sets (with and without shrapnel). It can be seen that the presence of the shrapnel changes the relationship between the cross and co-polar cross section.

Sample results for a subject with and without metallic shrapnel are shown in Figure 4. There is a substantial change in the relative distribution of received radar returns – in particular, higher returns are detected in the cross-polarised channel with the shrapnel present.

The final radar modality was detection of acoustically-induced vibrations using micro-Doppler measurements with millimetre-wave radar. Similar approaches have previously been proposed and experimentally demonstrated, for both people screening[3],[4] and remote biometric sensing.[5]

A loudspeaker is used to generate an audio signal, typically in the bass region (sub-300 Hz), and ideally below the frequency threshold of human perception. Small motions of an object in response to this audio create small perturbations (of phase and amplitude) in the reflected radar signal, which can be detected by the radar receiver if the motion is large enough and the radar is sufficiently sensitive.

Broadly speaking, this is a means of identifying the mass or inertia of objects within a bag or under clothing, together with their key radar properties. Small, low-mass objects with a large radar cross-section move more and generate stronger signals than large, heavy objects with a small radar cross-section. The frequency response at different audio frequencies is also important.

This signature has yielded promising results in testing on static subjects, but has yet to be implemented in the real-time prototype. This is due to the challenges of extracting these signatures in the presence of the much larger Doppler shifts caused by large-scale subject motion, in particular swinging arms and legs.

2.4 Ultrasound sensing

Ultrasound in the range 20–60 kHz is used to make pulse/echo ranging measurements of the bag. This frequency range was chosen as it is high enough to be inaudible to most people, but low enough that loss in transmission through bag materials is still tolerable. Ultrasound has previously been used for people screening applications,[6] but, to our knowledge, remote sensing of bag contents is a relatively novel application.

The reflected signal from each bag and its contents varies dramatically, depending on the density, porosity and acoustic impedance of each layer and object within the bag (Figure 5). We found that multiple microphones were required to sample at different heights. Ultrasound is very good at identifying (and thus clearing) simple, mostly-empty, thin-layered bags. This sensor acts on different physical properties and provides complementary information to the wide-band, millimetre-wave radar.

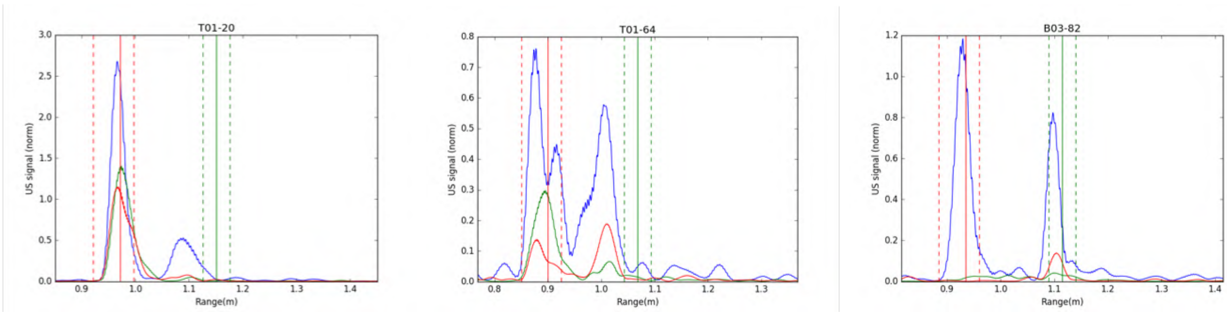

Figure 5. Ultrasound reflections from bags with different contents: plastic explosive simulant, powder explosive simulant and a mostly-empty bag (left to right). Note change in vertical scale between graphs. Each colour represents a different frequency of ultrasound (blue = 20 kHz, red = 30 kHz and green = 40 kHz).

We decided to operate in the 25–35 kHz region, following scoping studies, as this gives a good trade-off between directionality and penetration of most bag materials.

2.5 Data processing

The detection approach used in AcES is built around the extraction and combination of a number of features from the raw sensor data. The features are parameters representative of characteristics of threat objects and bags or of benign bags and their contents. Typical features include the dimensions and volume of the bag, magnitudes of different sensor responses, timing of radar returns and ratios of responses such as co/cross-polarised return. Overall, some 50–100 features have been identified based on interactions between the various types of sensing used to interrogate the subject.

This concept differs from the approach used in many “deep learning” applications, where a deep neural network is often fed the raw sensor data. This data is used without much filtering, so that the algorithm can directly infer features from the unprocessed information. We have not chosen to follow this route, as it requires orders of magnitude more data than we could feasibly provide. In addition, we believe the ability to incorporate prior knowledge of the application and sensor physics into our model makes our current approach much more suited to the task at hand.

Statistical classification and machine learning techniques determine the most relevant features and their combinations to some extent during the training process, but we have used our knowledge of physics and our model of threat and benign bag characteristics in our “feature engineering”. In short, we have selected features that are known to represent important characteristics of threat or benign items and thus are particularly relevant.

Four examples of these engineered features are provided below.

- The ratios of co- and cross-polarised peak radar returns, which are strongly indicative of the presence of highly-reflective, conductive, multi-reflecting and scattering materials. These materials include metallic fragmentation or complex metal objects, such as pipe bombs or large guns.
- Average radar (microwave or millimetre wave) return in a windowed region of range between the back of the torso and 30 cm behind the torso. This is likely to be higher for bags filled with objects with a high dielectric constant (such as plastic explosive).
- Physical volume of the bag as determined by the 3D camera, and its associated volume as measured with the 3D microwave imaging radar.
- Number and amplitude of distinct ultrasound peaks within a windowed region of range, indicating the number of strong scattering surfaces within a bag or directly under a clothing layer.

The informed choice of these features is a way of “feeding” knowledge into the algorithms and helping increase the signal-to-noise ratio (SNR) of the data used to make detection decisions. In addition, we have also created features capturing aspects of the sensor data that do not have an obvious direct physical interpretation, but could still be useful to a classification algorithm. We do not expect that all features will be relevant (indeed, many will be highly correlated), but the process of classifier fitting automatically selects useful features over weaker, less-significant features, which have no discrimination capability.

Several machine-learning approaches to the classification of the feature vectors were explored, alongside more traditional statistical clustering techniques. Experiments showed that random forests[7] gave good overall performance. This method has several advantages, including good resilience to overfitting, the ability to produce understandable outcomes and limited necessity for fine tuning a large number of model parameters to achieve good results. Neural networks were trialled in parallel and did perform better in some cases, but were much less consistent. These techniques also remove the ability to probe the reasoning for a detection verdict.

Data processing, feature extraction and classification take place in real time in several concurrently running software modules.

2.6 System integration

Various small-scale benchtop prototypes were produced throughout the development programme to enable studies of individual sensors and limited combinations. A “proof of concept” (PoC) prototype was developed in the most recent phase of the programme, which integrated all the sensors, power supply, computing, networking and other ancillary components into a single box (Figure 6).

We took a modular approach to designing the integrated system, rather than trying to minimise the size of the final system. One disadvantage of this approach is that the PoC does not represent the ultimate minimum form factor of the system. However, it had the significant advantage that the system was much easier to build, debug and customise. For example, future replacement of a radar only requires redesign of that module, rather than the whole system. Similarly, extra modules can be added or removed without compromising the form factor.

Figure 6. Initial AcES PoC prototype

We focussed on ensuring that each sensor ran as a self-sustaining, fault-tolerant and independent subsystem. Most of the sensor subsystems contain a small, dedicated computing module (such as a Raspberry Pi) to carry out some portion of the data processing and data reduction, thus reducing demand on the central host computer. Each sensor module thus only requires a single DC power connection and a standard Ethernet connection to the central processing hub.

Data collation from the multiple sensor modules, data processing, feature extraction and classification are carried out in real time on a small form-factor PC inside the main system box. The system screens people and their bags in motion in real time, delivering “safe or threat” verdicts several times per second with negligible latency. These results can be viewed on an external system (such as a laptop or tablet) and we have also demonstrated integration with an external CCTV system, with the AcES verdicts shown as an overlay on the video output.

Two identical systems were built so that the front and back of a subject could be screened simultaneously. The two systems were deployed at opposite corners of a $\sim$1 m wide and 4 m long channel during data collection and trials, representing the current maximum working range of the systems (Figure 7). This range reduces the risk of sensor crosstalk and ensures subjects are screened at a reasonably shallow angle (less than 15°), so that they remain in the workable field of view of the system for sufficient time. It should be noted that the systems could also be run back-to-back (screening in opposite directions), with both integrated into a single bollard structure. This further minimises any crosstalk issues, but removes the ability to screen the front and back of a subject simultaneously (rather than sequentially). It is theoretically possible to screen up to 2000 people per hour with a single pair of systems, allowing for appropriate gaps between people (at least 2 m to minimise occlusion). The PoC system was demonstrated during a field exercise in which two systems were left unmanned and operating autonomously for several days, successfully gathering data from several thousand subjects.

3. RESULTS

Over 2000 separate runs from a wide range of subjects, bags, benign contents and threat contents were collected during testing and trialling of the system, producing a data set that was used to optimise, train and test the

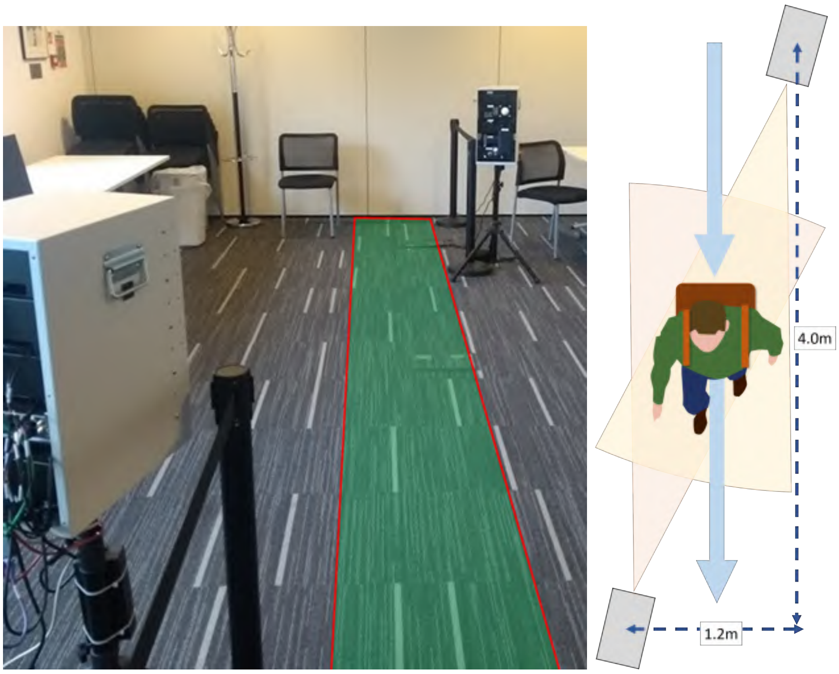

Figure 7. Photo of system set-up with walking path highlighted in green (left) and diagram of this preferred layout showing a person being screened (right).

detection algorithm.

We created a test set of 60 benign bags covering a wide range of sizes, contents, masses, homogeneities and metal contents (designed to represent a variety of venues and application areas), and a similarly varied set of threat bags representing realistic and credible threats. A range of threats was also designed and selected for concealment under clothing, including PBIEDs and large firearms. Subjects varied their outer clothing (coats) and benign pocket clutter (wallets, mobile phones, keys, etc.) between runs to ensure that the base level of complexity represented the true detection challenge faced in real high-footfall screening applications.

Results for subjects with and without bags are shown in Figure 8. We followed best practice in developing classifiers by divided our data into training and testing sets. All detection performance results in the following sections are reported solely from sets of unseen "hold-out" testing data, which the classification algorithms had not previously been exposed to during the training phase. Testing data consisted of completely novel runs that the algorithm had not "seen" before, although specific threat items, bags and subjects were included in different combinations in the training data.

Our initial results demonstrate that detection performance is consistently better than chance. This held true even when we presented previously unseen subjects or unseen bags in the testing data (representing a stricter hold-out condition than the data shown in Figure 8). Analysis of the classifier showed that all four sensors contributed to detection capability, demonstrating the value of the fused multi-sensor approach. Much better performance is seen for person-borne threats concealed under clothing when subjects are not carrying benign bags. This confirms the widely-accepted belief that divested person screening represents an easier challenge than combined people and bag screening.

We also created an additional, completely independent set of benign and threat bags for standalone validation tests, to provide independent confirmation of detection capability. Detection performance reduced when scoring these completely novel threat objects, although it remained significantly better than chance. This indicates that the system cannot currently completely generalise the detection problem or identify the optimal signature set. It is our view that its current performance is significantly hampered by the quantity of data available for classifier training.

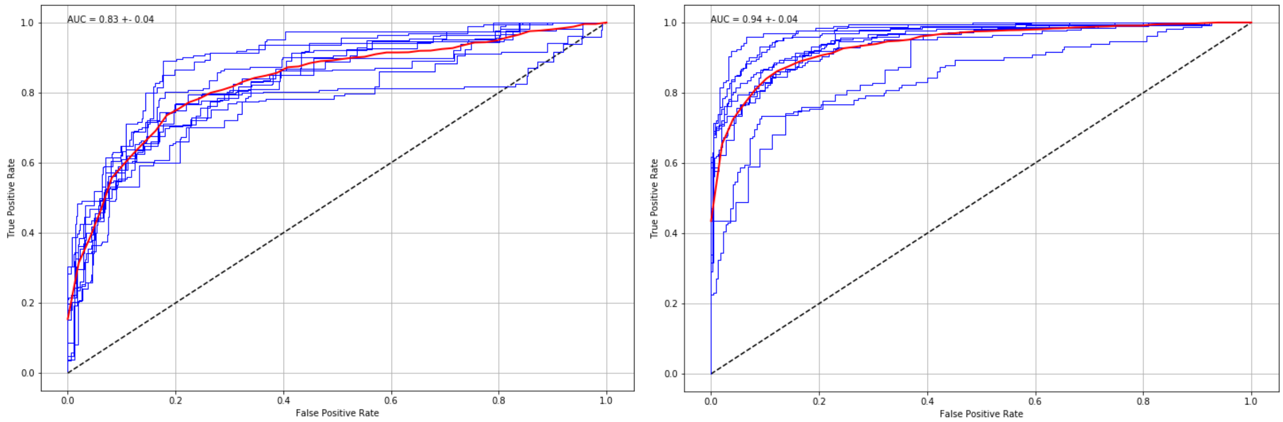

Figure 8. Receiver operating characteristic (ROC) curves for system performance on unseen data for subjects carrying either benign or threat bags (left) and subjects without bags or with person-borne threats (right). Each blue curve is performance on a specific hold-out set and the red curves represent the average of 10 different hold-out sets.

The quantity of data we were able to produce within the scope of the project was relatively small compared to many machine-learning applications, with less than 100 different bags and even fewer people. Additional data, especially with a more diverse range of threat bags and objects, would almost certainly improve the algorithm performance.

4. DISCUSSION AND CONCLUSIONS

A single screening lane comprising a pair of AcES units (for front and back screening) has a potential throughput of up to 2000 people per hour. This represents a substantial improvement over manual search, where a typical people and bag screening process clears only 300–600 patrons per hour per lane. A pair of AcES systems could therefore replace up to five manual search lanes. Our final false positive rate (FPR) of 20% would require a single search lane to resolve the 400 alarms per hour raised at peak throughput, but further optimisation would likely reduce this further. Alternatively, secondary screening technologies could be used to resolve alarms, such as metal detection and X-ray. In particular, AcES and our high-throughput X-ray system[8] are potentially complementary technologies.

Overall, our long-term vision that AcES could automatically clear between 75% and 90% of patrons at high-footfall venues seems eminently achievable. This represents a security solution that could deliver significant cost savings, as well as significantly reducing delays and inconvenience for customers. The results from this project emphasise that the AcES concept is a viable prospect, both from a technical detection perspective and operationally as a usable tool in high-footfall venues. Development work so far has resulted in a set of real-time, integrated and automated prototypes, which have been demonstrated successfully in real-world situations. The next stage is to further improve detection performance and demonstrate the technology in real-world field deployments.

ACKNOWLEDGMENTS

This project was funded under the Innovative Research Call in Explosives and Weapons Detection. This is a Cross-Government-funded programme sponsored by a number of departments and agencies under the UK Government's CONTEST strategy in partnership with the US Department of Homeland Security, Science and Technology Directorate. The views expressed in this publication are those of the authors and not necessarily those of funding contributors.

REFERENCES

[1] McMakin, D. L., Keller, P. E., Sheen, D. M. and Hall, T. E., "Dual-surface dielectric depth detector for holographic millimeter-wave security scanners," Proc. SPIE **7309**, Passive Millimeter-Wave Imaging Technology XII, 73090G (2009).

[2] Andrews, D. A., Harmer, S. W., Bowring, N. J., Rezgui, N. D. and Southgate, M. J., "Active Millimeter Wave Sensor for Standoff Concealed Threat Detection," IEEE Sensors Journal **13**(12), 4948–4954 (2013).

[3] Buersgens, F., Acuna, G. and Kersting, R., "Terahertz imaging of concealed objects by acoustic phase detection," Proc. SPIE **6949**, Terahertz for Military and Security Applications VI, 694905 (2008).

[4] Redo-Sanchez, A., Kaur, G., Xi-Cheng Zhang, Buersgens, F. and Kersting, R., "2-D Acoustic Phase Imaging With Millimeter-Wave Radiation," IEEE Transactions on Microwave Theory and Techniques **57**(3), 589–593 (2009).

[5] Petkie, D., Bryan, E., Benton, C. and Rigling, B., "Millimeter-wave radar systems for biometric applications," Proc. SPIE **7485** (2009).

[6] Felber, F. S., Davis III, H. T., Mallon, C. E. and Wild, N. C., "Fusion of radar and ultrasound sensors for concealed weapons detection," Aerospace/Defense Sensing and Controls, 514–521, International Society for Optics and Photonics (1996).

[7] Breiman, L., "Random Forests," Machine Learning **45**(1), 5–32 (2001).

[8] Kemp, M. C., Pollock, S., Crick, D.R., and Winter, L.J., "Fast, automatic, low-cost X-ray bag screening for mass-casualty threats," Proc. SPIE **11738**, Anomaly Detection and Imaging with X-Rays (ADIX) VI, 1173818 (2021).