Biometric authentication in the real world

Dr Tony Mansfield from the National Physical Laboratory reports on real life performance of biometric systems…

Over recent years there has been considerable growth in interest in the use of biometric systems for personal authentication. By using biometrics, authentication is directly linked to the person, rather than their token or password.

On the basis of media hype, you might conclude that biometrics will provide 100 percent identification. But what if actual performance is far less impressive? How and where biometric systems are deployed will depend on their real-life performance, which must therefore be appropriately measured.

The UK Biometrics Working Group (BWG) co-ordinates the Communications Electronics Security Group (CESG) Biometrics Programme, the goal of which is to enable the use of biometric authentication in Information Age Government. To achieve this, the BWG is establishing the security credentials of biometric technology through the development of test standards and protocols, and providing security assurance through International Common Criteria evaluation and certification.

As part of this work, CESG sponsored the National Physical Laboratory to carry out a test programme evaluating some of the leading biometric technologies. The first objective was to provide factual, vendor-independent data on the performance of biometric devices. Such information on the general capability of biometric technology will help in the development of policy on how and where biometrics might be used within Government and elsewhere. Further objectives were to validate the BWG proposed methodology for biometric testing, to support the development of the methodology for use with Common Criteria evaluations of biometric products and systems, and to act as a stimulus to later evaluations.

The systems selected for testing represented most of the common types of biometric technologies:

- Face — Visionics – FaceIt
- Fingerprint — VeriTouch – vr-3(U), using Infineon FingerTIP chip sensor
- Hand — Recognition Systems – HandKey II
- Iris — Iridian Technologies – IriScan system 2200
- Voice — OTG – SecurPBX Demonstration System

Also tested were an optical sensor fingerprint system and a prototype vein pattern system. The results for these two systems were less representative of the technologies, and are not shown here.

The evaluation scenario was 'access control' in a 'normal office environment'. This is a fairly typical application for biometric systems, most of which should perform close to optimally in such conditions. Users were to be co-operative, though relatively unfamiliar with using the biometric systems, as would be the case with infrequent users, for example.

The tests were conducted with 200 volunteers. These were drawn mainly from NPL staff, so there were no children, also women and older age groups were slightly under-represented. A further factor is that the volunteers generally had a positive outlook to technology; this might also influence performance.

All subjects attempted enrolment on each device, with repeat enrolment attempts if required. The only 'failures to enrol' were on the fingerprint system, where two people (1 percent) could not enrol due to the very poor condition of their fingerprints, and the iris system, where one person (0.5 percent) could not enrol a blind eye. Of course, in a larger and fully representative sample, we would encounter people with other

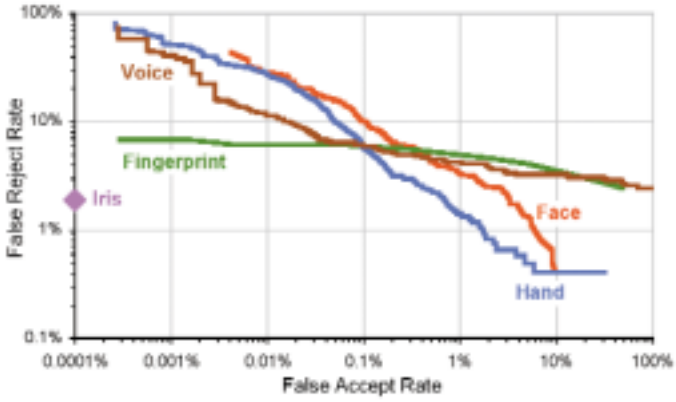

Figure 1. Detection error trade-off: FAR vs FRR

enrolment-preventing disabilities such as amputees. So for most systems the enrolment failure rate is slightly above zero.

To analyse identification accuracy, at approximately four weeks and eight weeks after enrolment, volunteers made three verification attempts on each system. These attempts were logged, to allow the frequency of verification failures (false reject rate) to be calculated. By comparing each attempt against the enrolment templates of all other volunteers, we could also observe the frequency of false acceptance (false accept rate).

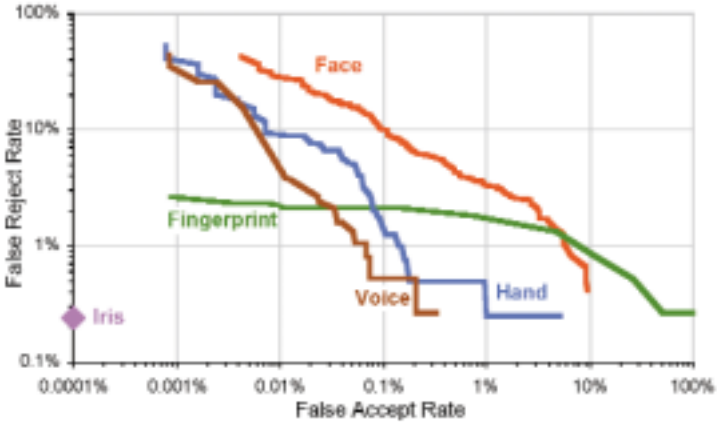

Figure 2. Detection error trade-off: Best of 3 attempts

Most biometric systems have an adjustable 'decision threshold' for a trade-off of usability, ie, false rejections against security, ie, false acceptances. The relationship between the false reject rate and false accept rate is best shown using 'detection error trade-off' curves as in figure 1. The lower and further left on the graph, the better the performance. It is seen that the iris system had the best accuracy, with 1.8 percent false rejections and no false matches in over two million comparisons. Of the other systems, fingerprint performs best for low false acceptance rates, while hand geometry can achieve low (below 1 percent) false rejection rates if false acceptance is not too critical. This illustrates that there is not a universally 'best' biometric system – the best system for high security may not be the best for high accessibility.

In their normal mode of operation, many systems allow multiple attempts. Figure 2 shows that for most systems, allowing three attempts considerably improves performance. In the case of face recognition, the improvement in false reject rate was offset by the worsening in false accept rate and no change is shown.

We also measured user throughput on each of the systems. This tends to be affected much more by system design, than by the performance of the image capture and matching algorithms. For example, the speed of voice verification is dictated by the user prompts. For high-throughput applications, the hand geometry system and optical fingerprint sensor were the fastest to use, capable of verifying up to eight people per minute. The other systems were slightly slower handling up to six verifications per minute.

When systems are used for identification, rather than verification, the input biometric must be compared against many enrolment templates, and throughput of the matching algorithm becomes important. The supplied algorithms processed from approximately one thousand comparisons per minute for voice and face recognition, up to 1.5 million comparisons per minute in the case of iris recognition. It is likely that each of the algorithms could be further optimised for one-to-many matching.

We also looked at performance differences attributable to the category of user. Looking at all seven systems it appeared that men and younger users were generally more likely to be successfully verified than women or older age groups. In the case of fingerprints age seems more significant than gender.

Our observations during the evaluation show that many of the false-rejections were due to user error caused by unfamiliarity with the system. Thus, were the systems being used on a daily basis as part of peoples' jobs, performance would be noticeably better. However this level of system familiarity is unlikely to be the case for many envisaged public applications. With such applications our results show that a fallback system will be essential to deal with the small proportion of genuine verification attempts falsely rejected.

The evaluation has provided factual, vendor independent data on the performance of biometric systems. The results are indicative of the general capabilities of current biometric systems in good conditions. With other environments, user population, types of application, or for systems other than those tested, performance figures are likely to be different.

The testing methodology proposed by the BWG has been shown to be feasible for evaluating performance of biometric systems at their current level of accuracy.

The results, and test report available at http://www.cesg.gov.uk/biometrics have generated a lot of interest in the biometrics community, renewing interest in performance testing and encouraging other companies to consider Best Practice compliant performance evaluation.

'How and where biometric systems are deployed will depend on their real-life performance, which must therefore be appropriately measured.'

NPL National Physical Laboratory
Dr Tony Mansfield
Centre for Mathematics and Scientific Computing
National Physical Laboratory
Queen's Road
Teddington
Middlesex TW11 0LW
Tel: 020 8943 7029
Fax: 020 8977 7091
tony.mansfield@npl.co.uk
www.npl.co.uk