Experiences from Large Scale Testing of Systems using Biometric Technologies

ERNCIP thematic group - Applied Biometrics for CIP

Deliverable: Experiences from Large Scale Testing of Systems using Biometric Technologies

Task 2

Peter Waggett, IBM (UK)

March 2015

The research leading to these results has received funding from the European Union as part of the European Reference Network for Critical Infrastructure Protection project.

Report EUR 27190 EN

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Georgios Giannopoulos
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 721, 21027 Ispra (VA), Italy
E-mail: erncip-office@jrc.ec.europa.eu
Tel.: +39 0332 78 6211
Fax: +39 0332 78 5469

http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC95455

EUR 27190 EN

ISBN 978-92-79-47445-3

ISSN 1831-9424

doi:10.2788/33813

Luxembourg: Publications Office of the European Union, 2015

Printed in Italy

Experiences from Large Scale Testing of Systems using Biometric Technologies

Table of Contents

Abstract

The intended readership of this paper is organizations looking to implement very large scale identification systems (e.g. national scale systems which may run to many millions of individuals). Many of the lessons and issues identified will also be useful for organizations looking to develop more general systems based on biometric technology.

This paper describes a systematic approach to testing based on lessons learnt from a case study of large scale testing of biometric systems. This approach will enable the performance of the proposed biometric matching system to be characterized to ensure that it is 'fit for purpose' and that the benefits outlined in justifying the system can be met.

1. Introduction

Biometric technologies can be used in a variety of ways to provide benefits to the operator of a system which requires a more secure recognition of individuals. Examples of generic applications include:

1. Verifying that someone who has been previously enrolled in a system is the same individual who is claiming an identity. The core process is a 1 to 1 comparison of biometric characteristics collected at enrolment and recognition. Physical access control to buildings or rooms within them could make use of 1 to 1 systems.

2. Identifying an individual with biometric characteristics collected in a biometric reader as someone who is already in a data store of biometric characteristics, often with a link to biographic identities such as a name. This is often described as a 1 to many comparison of biometric data.

3. Confirming that an individual is not in a database which contains biometric data, or to de-duplicate a database with possible multiple instances of a person's identity.

The intended readership of this paper is organizations looking to implement large scale identification systems (e.g. national scale systems which may run to many millions of individuals). Many of the lessons and issues identified will also be useful for organizations looking to develop more general systems based on biometric technology.

This paper describes the lessons learnt from large scale testing of biometric systems, in particular of systems of type 2. One key principle it that it is good practice to decide on the testing approach early on in the design cycle so as to facilitate the testing programme. The testing discussed is needed as a part of the evaluation prior to any potential deployment.

1.1 The Necessity of Biometric Testing

During the planning of a system, operators may have a requirement to assess how well systems with biometric components might work in order to decide whether (or not) the accrued benefits justify the expenditure of time and money.

Having made the decision to go ahead, they could specify the minimum performance for an application in the tender for procurement of systems (which could include other, non-biometric methods of recognition of individuals). Operators could also require that integrators bidding for the tender demonstrate their capability of meeting these minimum performance levels. The integrator, in turn, is reliant on the supplier of the underlying biometric technology providing results on a product which, when integrated in a system, will satisfy such minimum requirements.

Finally, operators would want to be assured that the delivered system works to the tender specification through tests in an operational context.

All three actors, the supplier, the integrator and the operator (or any proxy used in procurement), need to agree on a common approach (or approaches) that will provide this assurance, even though each actor may have different aims, priorities and views on the best way to test the system.

This paper describes the lessons learnt from large scale testing of biometric systems, in particular of systems of type 2 above. One key principle it that is good practice to decide on the testing approach early on in the design cycle so as to facilitate the testing programme.

2 Testing Biometric Systems

Applications which use biometric technologies are many and varied. For systems deployed at a large scale, the cost of failure can run to tens or hundreds of millions of pounds as well as the associated loss of reputation. And replacement systems may require users to readjust to news ways of working.

Therefore, a lot rides on making sure that the application performs to the desired specification – and not just that the biometric subsystem functions correctly. However, without a properly functioning biometric subsystem, it is very unlikely that the benefits of the deployed application will be realized! (See the schematic below which shows the relationship between the various terms.)

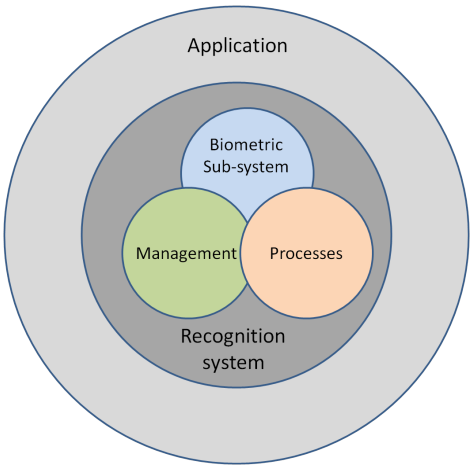

Fig 1 Relationship between a biometric sub- system and other components for an application[1]

One of the two fundamental aspects of the operation of a biometric system is the automated comparison of biometric characteristics to obtain a similarity score. If the similarity score meets (or exceeds) a predefined threshold, the characteristics are deemed to be matched, and the individual is recognized.

The other fundamental aspect acknowledges the limitations of the human-technology interaction which impacts on the representation of individuals' identities by templates in the system; the collection and processing of a biometric characteristic will differ each time the user is involved in the biometric process. For example, a finger may have been pressed unevenly on the sensor's surface, the subject may have been smiling or facing away from the camera (whereas on enrolment they will have been asked to face forward and not smile).

In performance testing of biometric applications, we seek to measure the likely rates of failure:

- Failure to Enroll (FTE). For a variety of reasons it may not be possible to capture biometric characteristics from an individual in the first place. This is not the fault of the individual and they should not be stigmatized or disadvantaged from using the system.
- False Non Match. This is an incorrect failure to identify an individual, who is not matched by the system against their own enrolled record.

[1] From draft of ISO/IEC 30124, *Code of practice for the implementation of a biometric system*

- False Match. This is an incorrect identification of an individual through matching their biometric characteristics to an existing enrolled record of someone else
- Failure to Acquire. For a variety of reasons it may not be possible to capture biometric characteristics to match to existing enrolled biometric characteristics. This may not be the fault of the individual and if it is not then they should not be stigmatized or disadvantaged from using the system. If this persists, the individual may need to be enrolled again into the application.

The corresponding performance metrics are the rates at which these errors occur in the test population: FTE, False Non Match Rate (FNMR), False Match Rate (FMR) and FTA. [2]

Other terms and definitions can be found in the ISO Biometric Vocabulary standard. [3]

In addition, there will be metrics specifically associated with tests addressing IT security requirements and assessing the usability of the system by the intended population.

2.1 Why does Testing of Biometrics Systems Need to be undertaken on a Large Scale?

Where the application is intended to identify individuals on a scale of millions, it has to work reliably under the widest range of operational conditions and for a wide diversity of people. That is, the biometric subsystem is required to work at the desired operating point for metrics such as FNMR, FMR and FTE and associated security and usability parameters.

The statistical basis of biometric comparison is affected by a number of factors. These include,

- the underlying population demographics (e.g. age, gender, ethnicity and disability)
- the size of the database, and
- the performance of all elements of the biometric subsystem used to acquire images and compare biometric characteristics.

There is unlikely to be a simple or calculable interdependence between these factors and the resulting performance of the system. This means that although test protocols can be guided by previously available test results on similar systems, the performance for a specific application will have to be determined by a bespoke testing approach. Publicly available test results on similar systems also require careful consideration for a number of reasons, as outlined in the next section.

The reader interested in more background on biometric testing is referred to the Best Practice Guide by Mansfield and Wayman (2002) which includes rules of thumb around the optimum sizes needed for test data sets[4]. ISO standards provide further information[5].

[2] When 1:1 verification systems are tested, the metrics should be considered on a *per transaction* basis; several attempts are allowed before a transaction is deemed to be completed, In this case, the error rates are presented as False Rejection Rate, FRR, and False Acceptance Rate, FAR, respectively.

[3] For a freely available copy of ISO/IEC Biometric Vocabulary 2382-37:2012, see: http://standards.iso.org/ittf/PubliclyAvailableStandards/index.html.

[4] Mansfield, A.J. and Wayman J.L., (2002), *Best Practice Guide for Biometric Testing*, NPL Report CMSC 14/02

2.2 What are the Issues Associated with Use of Publicly Available Test Results?

Publicly available test results are available from the following sources:

- Vendors of biometric equipment
- Open testing performed by publicly funded bodies (e.g. the National Physical Laboratory in the UK or NIST in the USA)
- 'Benchmark' testing conducted by organizations procuring biometric systems
- Academic led testing for research.

Although they can provide valuable input to any evaluation of biometric technologies, and indeed may have to be used at early stages of any procurement programme for justifying business decisions, they all suffer from a number of limitations that mean that they cannot be taken to be definitive - or even an adequate basis - for decisions on the procurement of large scale biometric systems. Where these are used in the initial definition phase, subsequent testing should be used to revisit the initial conclusions, thereby ensuring that these are still valid, as well as confirming the validity of any business decisions based on initial test results.

The principal limitations on the use of such publicly available results include:

- Similarity of application. Seemingly small changes in the system design and context of operation can affect the transferability of the results of tests on one system to the one under consideration by the operator. If there is no cost-effective alternative, and as a minimum, a careful analysis of differences in demographics, conditions of collection, etc. between the two systems should be undertaken, with tests commissioned to explore the impact of noted differences.

- Progress in biometric technologies. Results obtained in earlier tests will inevitably have used older algorithms, and hence may provide an unduly pessimistic assessment of current performance.

- Availability of test data sets of the required size. Biometric data is rightly regarded as personal data and its use in testing will require significant care to ensure that all of the pre-agreed protocols around access to, and the use and deletion of, such data are followed. These tight access restrictions result in very few large scale test data sets being available for other groups against which to test new systems.

- Representative test data sets. As discussed in the previous section, experience shows that the performance of biometric systems is dependent on the specific demographic makeup of the population. It is therefore important that all test data sets are representative of the target population. Although synthetically generated test data sets have been investigated, the current advice is not to rely on these until further research demonstrates their robustness.

---

[5] ISO/IEC 19795-1:2006 Biometric performance testing and reporting -- Part 1: Principles and framework. For operational testing (Type E above), please refer to ISO/IEC 19795-6:2012 Biometric performance testing and reporting -- Part 6: Testing methodologies for operational evaluation

- Confidentiality of test results. There is a reluctance to disclose details of the test results by operators for fear of inadvertently releasing information which could be of use to those aiming to subvert the biometric subsystem or any applications which make use of it.

- Commercial and competitive pressures. Vendors of biometric technology are in a very competitive marketplace and hence are unwilling to disclose results which could position their products in a bad light.

- Costs of testing. The significant costs associated with large scale testing may mean that large scale tests may only cover a subset of the available technology and may not include the most appropriate technology for an application.

- Extrapolation techniques. Techniques and methods for extrapolating performance figures for large systems from small scale tests have been used in the past, but reliance on these extrapolations poses considerable concerns for the test house undertaking the test and the operator relying on the interpretation of results. Clearly, the larger the test, the less the need for such extrapolation with a resulting greater confidence in the prediction of performance on operational systems.

3 Approach to Testing

To illustrate the benefits of a structured approach to testing we now discuss a case study of a very large biometric matching system. Diverse biometric technology vendors participated in the programme and their insights have also been included in this section. The main lessons learnt centre on:

- Establishment of common ground rules on system design and testing for all participants
- Access to, and protective measures for, the very large datasets required for testing
- Use of specialist test facilities
- Understanding the test lifecycle.

3.1 Ground Rules

The testing needed for any large scale programme is likely to require a significant effort in terms of time, money and effort from all concerned and it is vital that a set of ground rules are jointly developed and agreed prior to the commencement of the programme to ensure that expectations of all stakeholders are surfaced and can be met.

Representative ground rules included:

- Specific requirements on the biometric sub system need to be clarified. For this system there was a requirement for a 'software' only solution. No specialized hardware was allowed, and the software had to be architected so as to provide a simple biometric comparison and matching service as a part of a Service Oriented Architecture.

- The system under test has to be representative of the system required by the operator. This seems obvious – but, for example, some systems can be configured to deliver either greater accuracy or a higher throughput – and test system design needs to reflect the operating point specified by the operator.

- Fixed and final timescales for the testing need to be agreed in advance to ensure that the overall testing timescale can be met. This can also provide an indication to the test house (or integrator) of how well the suppliers of biometric technology can deliver to time.

- The extent of automation (and use of ancillary data) in the interpretation of results needs to be defined in advance. In the case study, all interpretation was automated with results obtained directly from the system. In effect this was a 'lights out' operational context. In addition, testing was performed solely on the biometric data, and no use of metadata (e.g. age and gender) was allowed.

- The approach to failures to enrol (FTE) into the system needs to be defined in advance. In this case study, suppliers were not allowed to declare a failure to enrol for any of the supplied set of biometric characteristics. Any images which could not be processed into the system (for example because of an inability to identify critical aspects of the image) were counted as failures in identification. In this way the suppliers of the biometric subsystem ensured that the test results used the maximum coverage of test data.

- The approach to working conditions and access to equipment and data needs to be defined and agreed in advance, as these may impose constraints on personnel involved in the testing in meeting their deadlines.

- Finally, the basis on which testing will be evaluated and recommendations reported on the preferred technology needs to be agreed in advance.

3.2 Test Data

The test data needed for any testing programme needs to be agreed in advance, and it should be representative of that to be encountered in the operational solution. Both training and test data should be provided in a timely fashion by either the operator or a third party so as not to delay the test schedules.

The security and tracking of all data is likely to be highly constrained and its management will require significant investment in terms of time, money, facilities and effort.

3.3 Test Facilities

The hardware and facilities needed to test biometric subsystems successfully at scale are likely to be complex and expensive to install, run and maintain. They are likely to be dedicated to run the specific tests, to ensure that throughput results are representative of the operational conditions with all test data capable of being monitored and controlled.

There may be a requirement to destroy some elements after testing (e.g. disk drives) to ensure that no personal or other sensitive data remain undeleted.

3.4 Testing Programme

Our experience is that the testing programme can be quite complex, but will consist of at least four phases:

- Set up and Tuning Phase. This allows the suppliers of biometric technologies to set up their systems in a way that presents their product in the best possible way for subsequent phases of the testing. This phase requires access to both training and test data that is representative of the main test dataset, along with the correct results (for pairs of matched and non-matched images). This allows optimization of the performance of biometric matching algorithms for subsequent testing on the full dataset.

- Vendor Selection Phase. This is the main test phase and is targeted at selection of the optimum technology to be used in the subsequent phases of testing, as well as to select the technology to be included in the operational system. The aim is to obtain comparable results (and relative performance) from all suppliers so as to finalize a decision on the best system thereby avoiding unnecessary expense to individual suppliers. The most accurate matching technology may not be the 'best' technology, as it could also be the slowest or least affordable of the algorithms,

- Confirmation and Risk Reduction Phase. This final phase is only performed for the selected supplier of the biometric subsystem, but may include the running of additional 'blind tests' to confirm performance, and extensive manual analysis of each anomalous result so as to better quantify the overall system performance. The analysis of individual anomalous matches and missed matches will often highlight data entry errors in the test data provided to the test house.

- Operational Phase. Although outside of the scope of the case study, this there is a need to address the ongoing performance testing of the system. Planning for this testing should be based on the lessons learnt in the preceding phases.

4 Conclusions

A systematic approach to large scale testing can:

- characterize the performance of a biometric matching system to ensure that the operational application is 'fit for purpose' and that the benefits foreseen in the business case can be met.
- provide a reliable assessment of the performance of the operational system.
- meet the aspirations of all of the stakeholders in developing the application, to ensure confidence in the system's performance while minimizing the costs of testing systems in terms of time, money and effort.

European Commission

EUR 27190 EN – Joint Research Centre – Institute for the Protection and Security of the Citizen

Title: Experiences from Large Scale Testing of Systems using Biometric Technologies
ERNCIP thematic group - Applied Biometrics for CIP

Author: Peter Waggett, IBM (UK)

2015 – 16 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424

ISBN 978-92-79-47445-3

doi:10.2788/33813

Abstract

The intended readership of this paper is organizations looking to implement very large scale identification systems (e.g. national scale systems which may run to many millions of individuals). Many of the lessons and issues identified will also be useful for organizations looking to develop more general systems based on biometric technology. This paper describes a systematic approach to testing based on lessons learnt from a case study of large scale testing of biometric systems. This approach will enable the performance of the proposed biometric matching system to be characterized to ensure that it is 'fit for purpose' and that the benefits outlined in justifying the system can be met.

LB-NA-27190-EN-N

JRC Mission

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new methods, tools and standards, and sharing its know-how with the Member States, the scientific community and international partners.

Serving society
Stimulating innovation
Supporting legislation

doi:10.2788/33813

ISBN 978-92-79-47445-3

Publications Office