

Project no.: IST-2005-027351
 Project acronym: MTIT
 Project title: Minutiae template interoperability testing
 Instrument: Specific targeted research project
 Thematic Priority: Information Society Technologies
 Start date of project: November 1, 2005
 Duration: 17 months

**Deliverable D6.1:
 Methodology and processes for Minutiae Interoperability Testing
 including D 5.2:
 Certifying interoperability with already tested products**

Version 1.0

Due date of deliverable: February 28, 2007
 Actual submission date: March 31, 2007
 Lead contractor for this deliverable: NPL Management Ltd

Project co-funded by the European Commission within the Sixth Framework Programme (2002–2006)		
Dissemination Level		
PU	Public	✓
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Contents

- 1 Purpose of the document 1**
- 2 Recommended interoperability measures..... 1**
- 3 Data to be provided 2**
- 4 Phases of testing..... 2**
- 5 Adding new systems 3**
- 6 Analysis methods 4**
- 7 Examples 4**
- References 6**

1 Purpose of the document

Within the MTIT Project, interoperability testing has been conducted on minutiae encoding and comparison algorithms of the four vendor partners: Cogent, Motorola, NEC and Sagem. These tests have established the level of interoperability between these systems. These interoperable performance results refer only to the systems tested, and we should to consider (i) how to proceed to show that other implementations interoperate with this set at a satisfactory performance level; and (ii) a general methodology for showing interoperability of different sets of algorithms for the purposes of certification or to select the most interoperable algorithms in different usage scenarios (including scenarios other than those examined in this project).

The purpose of this report is to provide a method whereby minutiae algorithms of a further supplier may be tested to be shown to be equally interoperable with the systems already tested. The report therefore proposes a process for testing interoperability with the existing set of interoperating systems, specifying the database(s) to be used for testing, and the metrics that would be employed in judging interoperability. This report also describes a general methodology that may be used for future interoperability tests. Specific details of this methodology, dealing with the fingerprint minutiae interoperability addressed within the MTIT project, are provided in a revision (version 1.1) of MTIT deliverable 3.2 (Specification of a test and certification protocol) [MTIT D3.2] and are not repeated in this document.

2 Recommended interoperability measures

In the scenarios examined in the project, interoperability is considered between the minutiae extractions subsystems used to enrol the reference template, and subsystems performing minutiae extraction and comparison of the claim fingerprint. Our discussion of interoperability metrics and methods continues to consider such a scenario. However, the metrics and methods generalise to more complex scenarios where we might need to consider interoperability between three (or more) types of subsystems (for example when minutiae extraction of the claim and reference fingerprints, and comparison are performed by algorithms of three different suppliers).

The project would recommend three distinct methods of assessing interoperability of a set of systems. These are described below, with detailed methodologies given in section 6. It is important to note that these are not the sole measures that characterise the interoperability of systems, however these are understandable robust metrics.

1. Measurement of interoperable performance.

Interoperable performance is measured by the average interoperable performance of a set of systems with some calculated deviation.

E.g. Claim comparison systems C_1, C_2, \dots, C_m interoperating with reference generation systems R_1, R_2, \dots, R_n provide an average interoperable performance μ (with standard deviation σ).

2. Certification of interoperable performance

Average interoperable performance may be certified to be no greater than some previously determined target level. The target performance level applicable to a specified dataset might be established by prior work – for example knowledge that some existing implementations interoperate at this level – or by a specified system performance requirement.

E.g. Claim comparison system C_i interoperating with reference generation systems R_1, R_2, \dots, R_n provides an average interoperable performance $\leq \mu$ (at an appropriate confidence level).

3. Selection of most interoperable subset

The most interoperable set might be determined via excluding those claim-reference pairs that have performance that is significantly higher than the average interoperable performance of the whole set.

E.g. Claim-reference pairing $C_i R_j$ gives interoperating significantly higher than the average of interoperable subsystems (at an appropriate confidence level).

3 Data to be provided

Upon wishing to have a system tested the supplier would be provided with the following information:

- o The specification for the algorithms to be tested, in order that these will interface to the test tool [MTIT D4.4].
- o The specification of the minutiae templates, in accordance with the relevant parts of [ISO/IEC 19794-2].
- o (If available) a set of sample images taken from an example dataset, and the corresponding minutiae templates as encoded by the existing interoperable set of vendors. In the absence of a reference system and specification, this aids the supplier in knowing what they will be required to interoperate with. If such a data set is not available, it should be generated in an initial phase of testing;

4 Phases of testing

If no previous interoperable set of systems had been defined then an initial investigation phase shall be performed. This would evaluate the conformance to the minutiae template standards and to the requirements of the test tool interface and provide a baseline performance from

which improvements may be evaluated. A subset of finger images and templates from this phase would be distributed to all vendors to allow improvements in algorithms before the certification phase. Data generated in this phase would be retained for future use. This phase of testing should also allow the setting of an achievable interoperable performance criterion.

Having allowed vendors to update the systems they wish to be tested a second phase of testing, a certification phase, will be undertaken. This will be carried out on sequestered data, which should be appropriately selected in accordance with the guidelines in [ISO/IEC 19795-1], e.g. consideration of operational realism should be made. Interoperable performance matrices generated in accordance with [ISO/IEC 19795-4] (and as shown in the annexes of [MTIT D5.1]) will be generated.

At this phase of testing, conformance testing to the minutiae template standards is repeated on the new data. Any conformance error is counted as an (false non-match) error for that minutiae extraction system, but not as an error for a conformant matcher matching against a different conformant fingerprint record (i.e. no penalty is incurred for failing to match a non-conformant template).

The assessment of interoperability, as defined in section 2, should then be carried out on the results of the certification phase data. If available, tests on additional databases may be used to validate these findings.

5 Adding new systems

A new system wishing to be tested against an existing interoperable set as produced in section 4, would first be evaluated for conformance to the minutiae template standards and to the requirements of the test tool interface. The vendor would be provided with images from the investigation phase dataset (of the original interoperable set) together with corresponding templates generated by the interoperating systems.

A set of templates would be generated by the new system for the certification database. These would be used together with the corresponding minutiae templates from currently interoperating systems to run cross comparisons between minutiae templates that had not been previously compared by all matchers. This will combined with the previously obtained performance data to generate additional rows and columns extending the interoperable performance matrices.

The new system would be considered consistent with the interoperating set if the interoperable performance of claim-reference pairings involving the new system comply with measures 2 and 3 in section 2.

The number of systems with which a new system may be compared can rise as more interoperable systems are added. However in order to maintain a standard interoperability test, baseline average interoperable performance and standard deviation will be fixed as calculated for the original set of interoperable systems.

6 Analysis methods

This section provides detail to calculating the measures described in section 2.

1. The average interoperable performance (μ_I) is defined as the mean performance of all claim-reference pairs where the claim and reference are from different systems, the interoperable deviation (σ_I) is the standard deviation of the performance of those pairs.
2. To certify a new system's average claim performance (μ_C) as equal to, or better than, the target performance (p) at a 95% confidence level, we proceed as follows. If there are t genuine trials per claim-reference pair, and n reference generators in the interoperable set, then calculation of μ_C is based on nt trials. Treating these as independent, identically distributed means that the number of observed errors ($nt\mu_C$) follows a binomial distribution. If v is the maximum value for which the probability of v or fewer errors in nt trials when the error rate exceeds p is less than 5%, then for certification we require that $\mu_C \leq v/(nt)$. The critical value v may be computed by statistical packages (e.g. in Excel $v = \text{critbinom}(nt, p, 5\%) - 1$) or using the normal approximation to the binomial distribution. A new system's average reference performance (μ_R) can be certified in a similar manner.
3. A claim-reference pair will be considered an outlier to be potentially excluded from the interoperable set if the performance of that pair (μ_{CR}) is greater than μ_I with a 95% confidence level. That is, if $\mu_{CR} > \mu_I + \sigma_I z_{0.95}$ where $z_{0.95}$ is the 95th percentile of the standard normal distribution (i.e. the area under the tail of the standard normal distribution from $z_{0.95}$ to ∞ is 0.05).

7 Examples

In order to assess a new system with reference generation subsystem R_5 and claim comparison system C_5 against the interoperable set of subsystems ($R_1, R_2, R_3, R_4, C_1, C_2, C_3, C_4$) developed for the certification phase of the MTIT project the following datasets and performance values may be used.

- Data from DB_A2 would be released to the supplier to aid them in understanding the requirements for interoperability.
- DB_A1, which is sequestered at NPL, would be used for the certification phase of the testing. There is the possibility of using further databases to perform validation of the results (although permission to utilise the supplier databases, DB_S, DB_M and DB_C, expires at the end of the project).
- With DB_A1, the tested implementations gave an average interoperability performance of 1.8% (with a standard deviation of 0.4%) at a false match rate of 0.1%

- To demonstrate an interoperable performance of 1.8% or better with 95% confidence, noting that DB_A1 comprises 1066 genuine trials, the critical value for the number of observed false-nonmatch errors, below which R_5 or C_5 would be certified, is calculated to be 62. (E.g., in MS Excel, $62 = \text{critbinom}(4*1066, 1.8\%, 5\%) - 1$). So R_5 or C_5 would be certified interoperable if the observed FNMR, against (C_1, C_2, C_3, C_4) and (R_1, R_2, R_3, R_4) respectively, is at most 1.4% ($= 62/(4*1066)$).
- The criteria for a claim-reference pair to be considered an outlier at a 95% confidence level would be that the FNMR exceeds

$$\mu_1 + \sigma_1 z_{0.95} = 1.8\% + 0.4\% * 1.645 = 2.4\%$$

at a FMR of 0.1% on DB_A1. However, we note that the new reference encoder is tested in four claim-reference pairings, approximately quadrupling the chance of random effects showing an outlying interoperable performance (i.e. a 20% chance). Therefore, if a reference encoder is to be deemed non-interoperable on the basis of a single outlying claim-reference pair, we should require a higher confidence level. E.g. the reference generator might be excluded if it is part of an outlying claim-reference pair multiple times, or of an outlier at a 99% confidence level, i.e. has FNMR exceeding:

$$\mu_1 + \sigma_1 z_{0.99} = 1.8\% + 0.4\% * 2.33 = 2.7\%$$

at a FMR of 0.1%.

References

- [ISO/IEC 19794-2] International Standard ISO/IEC 19794-2:2005, Information technology – Biometric data interchange formats – Part 2: Finger minutiae data
- [ISO/IEC 19795-1] International Standard ISO/IEC 19795-1:2006, Information technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework
- [ISO/IEC 19794-4] ISO/IEC FCD19795-4:2007, Information technology – Biometric Performance Testing and Reporting – Part 4: Performance Interoperability Testing
- [MTIT D3.2] Specification of a test and certification protocol. (Public Report) Deliverable D3.2 of the MTIT Project
- [MTIT D4.4] Specification of a test interface. (Restricted to other programme participants – including the Commission Services) Deliverable D4.4 of the MTIT Project
- [MTIT D5.1] Interoperability testing report. (Restricted to other programme participants – including the Commission Services) Deliverable D5.1 of the MTIT Project