Discussions of Datasets in Biometric Standards
Michael Thieme
Novetta

NIST
Improving Biometric and Forensic Technology: The Future of Research Datasets
26-27January 2015

Overview

- What do ISO standards on biometric performance testing and reporting say about databases / datasets?

- Focus on ISO/IEC 19795-N: Information technology -- Biometric performance testing and reporting standards developed in SC37
  - Part 1: Principles and framework
  - Part 2: Testing methodologies for technology and scenario evaluation
  - Part 4: Interoperability performance testing
  - Part 5: Access control scenario and grading scheme

- What gaps could be closed through revisions or amendments to these standards?

NOVETTA From Complexity to Clarity.

ISO/IEC 19795-1: Principles and framework (1)

- General principals for biometric performance testing and reporting
- On test size
  - The size of an evaluation will affect how accurately error rates are measured
  - The number of people tested is more significant than the total number of attempts in determining test accuracy
- Sufficient samples shall be collected per test subject so that the total number of attempts exceeds that required by the Rule of 3 or Rule of 30 as appropriate
  - Rules such as the Rule of 3 and Rule of 30 provide lower bounds to the number of attempts needed for a given level of accuracy
  - However, these rules are over-optimistic, as they assume that error rates are due to a single source of variability, which is not generally the case with biometrics
- Collecting multiple samples on different days, or from different instances, can help reduce the dependencies between samples by the same person
- Uncertainty in the performance measures shall be estimated, determining whether the test was large enough
- The number and frequency of test transactions collected per test subject should be in keeping with the target application

NOVETTA

From Complexity to Clarity.

ISO/IEC 19795-1: Principles and framework (2)

- Test size and uncertainty are addressed in an informative annex
- Rule of 3, Rule of 30
- Number of comparisons to support a claimed error rate
- Variance of performance measures as a function of test size
- Estimates for variance of performance measures
  - Variance of observed false non-match rate
  - Variance of observed false match rate
- Estimating confidence intervals
  - Bootstrap estimates of the variance and confidence intervals
  - Subset sampling

NOVETTA

From Complexity to Clarity.

ISO/IEC 19795-2: Technology and Scenario Evaluation

- Methodologies for scenario testing (live subjects interacting with sensors) and technology testing (offline algorithm testing)
- Multimodal tests shall not use chimerical subjects
- A corpus may be considered "tainted" if a supplier…
  - a) has had possession of the corpus;
  - b) provided equipment used in collecting or processing the corpus
  - c) has previously been tested and tuned using the corpus.
  - When use of a tainted corpus is unavoidable, this shall be documented in the test report
- Data should not be used if supplier(s) have had possession of it
  - Previous testing / tuning of the system using the test corpus shall be documented
- Corpus validation: process whereby Test Subject data is screened to remove data not suitable for the purpose of the evaluation
  - Data is in the correct format, with correct instances, and with ground truth errors identified Experimenters shall report whether Test Subject data has been validated, including redaction methods and figures (if applicable)

NOVETTA

From Complexity to Clarity.

ISO/IEC 19795-4: Interoperability & Performance Testing

- Defines tests to address absolute performance, sufficiency, and interoperability available from biometric data formatted to comply with established standards, particularly in ISO/IEC 19794-N
  - MINEX, IREX, etc.
- To resolve the question as to whether the SIF is significantly inferior to proprietary formats …
  - The test designer should ensure availability of data sets large enough to resolve small differences in the chosen figures of merit
  - This test may use all of this material or not, depending on the measured performance and on formal estimates of the confidence intervals

NOVETTA From Complexity to Clarity.

ISO/IEC 19795-5: Access control scenario & grading (1)

- Addresses 'technical performance testing' of biometric systems and subsystems to be used for access control
  - Based on scenario testing – subjects in a facility executing real-time transactions
  - A framework for a general purpose test – "One size fits many (but not all)"
  - Centered on common access control application requirements
  - The supplier can submit to one testing process, and many potential customers can interpret the results for their application

- The only SC37 testing standard that (1) assigns grades based on error rates and (2) assigns grades based on error rates

Age						Gender	
<18	18-30	31-50	51-70	>70		Male	Female
0%	25-40%	25-40%	25-40%	0%		40-60%	40-60%

NOVETTA
From Complexity to Clarity.

ISO/IEC 19795-5: Access control scenario & grading (2)

- Min crew size: 230 subjects / 3450 impostor transactions
  - If 0 false acceptances occur, FAR is with 95% confidence not higher than 0.001 based on the "Rule of 3"
  - If false acceptances occur, then the FAR is above 0.001
- Test consists of 10 specific graded metrics
  - Transaction level error rates at three different levels of FAR
  - Attempt level error rates at three different levels of FAR
  - Determination of FTE
  - Verification transaction time at three different levels of FAR

NOVETTA **From Complexity to Clarity.**

Dataset-Related Gaps in ISO/IEC 19795-N Standards

- Guidance on datasets required to test emerging modalities
- Datasets comprised of or including artefacts (e.g. fake fingerprints)
- Video surveillance applications (standards in initial stages of development)

9

NOVETTA

From Complexity to Clarity.