M1/03-0405 - Proposal to Develop a New Standard: Biometric Sample Quality

1. Source of the Proposed Work Project

1.1. Title

Biometric Sample Quality Standard

1.2. Date Submitted

August 21, 2003

1.3. Proposer

M1

2. Process Description for the Proposed Project

2.1. Project Type

This is a Development Project (D).

2.2. Type of Document

This Development Project is expected to result in a Biometric Sample Quality Standard document that will specify a universal means by which to assess, quantify, express, and interpret the quality of a biometric sample. This standard would be structured modularly with a generally applicable main body that can be amended for use with different biometric types and applications at a future date as required.

2.3. Definition of Concepts and Special Terms

2.3.1. Existing Terms

This proposal uses vocabulary and terms as currently listed in M1-03/128: M1 Vocabulary and Definitions. Where more than one definition is provided for the same word, the Biometric Evaluation Methodology (BEM) v. 1 takes precedence.

2.3.2. New Terms

Following are new terms introduced in this proposal. Note that term definition is a goal of the proposed standard.

Attributes – characteristics of the subject of a biometric capture inherent in the subject itself that affect the quality of its biometrics sample

Cumulative Match Characteristic – A means to present results data of a detection task test, which plots the probability of identification (y-axis) vs rank (x-axis)

Detection Error Tradeoff – A means to present results data of a detection task test, plotting True Reject Rate to False Reject Rate

Effectiveness – the impact of an individual biometric sample on the overall performance of a biometric system

Effectiveness Score – An objective, quantitative assessment of the effectiveness of a biometric sample.

Environment – The physical surroundings and conditions where biometric capture occurs, including operational factors such as operator skill

Fidelity – The degree of similarity between a biometric sample and its subject

Fidelity Score – An objective, quantitative assessment of the fidelity of a biometric sample.

Improvability – The degree to which the effectiveness of a biometric sample can be improved upon through recapture.

Operator – The individual who processes a user in a biometric system, performing capture

Performance –An assessment of the TAR, TRR, FAR, FRR, and FER of a biometric system

Quality – An assessment of the predicted impact of a biometric sample on the performance of a biometric system

Quality Score – A normalized, objective quantification of either the fidelity of a biometric sample or the predicted effectiveness of a biometric sample on the performance of a biometric system. Such a score might be further specified to reflect either predicted FAR, FRR, or FER.

Sample – An interpretation of a physical human feature such as a fingerprint or facial image used for identification or verification using biometric techniques

Subject – The physical body part represented by a biometric sample

2.3.3. Concepts

The term "quality" as it is currently used in the field of biometrics has several connotations, depending on context. Two currently prevalent uses are to subjectively reflect:

1. the *fidelity* of a sample to its original subject. A quality score based on fidelity reflects the accuracy of a sample's representation of the original subject.
2. the *effectiveness* of a sample within a biometric system. The predicted or observed positive or negative contribution of the given sample to the overall performance of a biometric system, where performance reflects the FAR, FRR, and FER of a given system.

These two uses present two paradigms for calculating and assigning quantitative quality scores:

1. A "bottom-up" score reflecting fidelity, based only on the sample (image, pattern, and/or features) itself, and,
2. A "top-down" score reflecting predicted effectiveness. This score might attempt to a) exploit some correlative relationship between sample characteristics and system performance, or b) draw on the sample's performance in controlled tests.

Fidelity: Bottom-Up Approach

Bottom-up fidelity information is derived from the sample only. With no effectiveness prediction capability, this information is useful primarily if it can discern between different contributors to its effectiveness. That is, will the samp le exhibit poor effectiveness because of inherent attributes, or can it be improved upon if recaptured?

There are three factors in the creation of a biometric sample that affect its fidelity and effectiveness:

1. Equipment: Different equipment (hardware and software) will have different impact on sample effectiveness. Hardware factors include scanner resolution and bit depth, and software factors include the image compression ratio.
2. Subject: Different subjects will impact image quality based on inherent attributes of the subject, such as cracked or small fingerprints for which extraction and matching algorithms are not typically optimized (though future innovations can bring improvements here).
3. Environment: Different operational conditions will contribute differently to sample quality. These include skills and habits specific to a particular operator, or environmental conditions, such as lighting or temperature. For example, these can result in smudged, incomplete, or misplaced fingerprints, and could conceivably be improved upon with a reattempted capture.

A biometric sample will reflect some degree of each of these three factors, and the ability to identify and differentiate between these factors is valuable in improving a biometric system's performance. A sample

with high fidelity can exhibit poor effectiveness. For example, if a sample is given a poor effectiveness score upon capture, an operator would benefit from knowing whether or not a capture reattempt would have a reasonable probability of resulting in a significantly improved score. This might be referred to as the sample's "improvability". A sample with low effectiveness due to inherent attributes of the subject will demonstrate a low degree of improvability, and use of a different biometric may be advisable. Alternatively, a sample with low effectiveness but high improvability would indicate equipment or environmental problems that could be improved upon, and recapture is advisable.

Effectiveness: Top-Down Approach

A quality score using a top-down approach reflects predicted effectiveness of the sample in a biometric system. This score might be calculated using algorithms derived from correlative relationships between sample attributes and effectiveness. It might alternatively use quality scoring guidelines derived using standardized test biometrics and data.

Factors affecting biometric system performance fall into three categories:

1. Sample quality (fidelity and effectiveness)
    i. Equipment
    ii. Subject
    iii. Environment
2. Feature extraction
3. Matching algorithm

These three are interdependent, as changes in any one will impact the contribution to (or detraction from) system performance of another, so an effectiveness score is relevant for the given extraction/matching environment in which it operates. It is important to recognize that such a quality scoring system is not independent of the extraction and matching system being used to create these scores. This representation may be a multi-dimensional problem.

		Fidelity	
		Low	High
Effectiveness	Low	Low fidelity and low effectiveness indicate that sample effectiveness and system performance could be improved with a new sample with higher fidelity, so recapture is recommended.	High fidelity and low effectiveness indicate that sample effectiveness and system performance cannot be significantly improved with a new sample, and an additional biometric should be recommended.
	High	Null. Samples with poor fidelity demonstrate inconsistent effectiveness and reduce system performance. They typically will not demonstrate high effectiveness.	Samples with high fidelity and high effectiveness represent successful capture.

2.4. Expected Relationship with Approved Reference Models, Architectures, etc.

A final document is expected to be referenced by those data format specifications produced by M1.1 which provide fields for interchange of quality data.

2.5. Recommended INCITS Development Technical Committee

INCITS Technical Committee M1 Ad Hoc

2.6. Anticipated Frequency and Duration of Meetings

Quarterly meetings in conjunction with M1(.1) meetings, and quarterly telephone conference calls held between M1 meetings.

2.7. Target date for Initial Public Review (Milestone 4)

It is estimated that the draft standard document would be ready for submission to INCITS for public review within 12 months of initiation.

2.8. Estimated Useful Life of Standard

There is no known limitation on the useful life of the proposed standard.

3. Business Case for Developing the Proposed Standard

3.1. Description

The proposed standard would establish universal, objective specifications for assessing, quantifying, expressing, and interpreting biometric sample quality, differentiating between the fidelity and the effectiveness of the sample. The standard would contain several annexes, each pertaining to a particular type of biometric (ie. finger, face, iris). The variability of biometric sample quality requirements for different applications (ie. one-to-one matching, one-to-many matching) would also be considered.

3.2. Existing Practice and the Need for a Standard

There are several applications of biometric sample quality scoring. The proposed standard would, at a minimum, address the use of quality scores to provide an operator or system administrator an indication of the:

1. predicted performance of a particular sample within a biometric system (or similarly, the impact of the sample on the overall performance of the system)
2. potential for performance improvement achievable by recapturing the sample.

Through this feedback, the performance of a biometric system can be significantly improved by reducing the number of poor biometric samples submitted upon enrollment. In addition, operational efficiency can be aided by the ability of the system to assist an operator in the decision to a) accept the sample, b) reject the sample, or c) reattempt the capture.

There is currently no standard for the quantitative calculation and expression of biometric sample quality and its interchange. While BioAPI v 1.1 specifies a structure and provides guidelines for providing a quality score (see Appendix), a new standard will help to define a less subjective quality scoring system useful for some applications as described above.

3.3. Implementation Impacts of the Proposed Standard

3.3.1. Development Costs

Labor is expected to total approximately 12 staff months.

3.3.2. Impact on Existing or Potential Markets

It is expected that the availability of a universal quality metric system will ultimately improve the performance of biometric systems through improved data feedback in the operation of the system, and the improved design of these feedback systems.

3.3.3. Costs and Methods for Conformity Assessment

The cost and method for conformity assessment is not known at this time, but can be expected to be an integral part of this standard.

3.3.4. Return on Investment

There is no known data to make such an assessment. If we develop a biometric sample quality standard, it can be widely used for several industries, such as MRTD for biometric data, government smart card for face image, and ten-print for AFIS (civil and criminal). To make an interchange of biometric data, the quality is very useful and shall be used.

3.4. Legal Considerations

3.4.1. Patent Assertions

There are no currently known relevant patents.

3.4.2. Dissemination of the Standard

The drafts of this standard will be disseminated electronically on the INCITS M1 document register.

4. Related Standards Activity

4.1. Existing Standards

• ANSI/INCITS 358-2002 Information Technology – BioAPI Specification v 1.1

4.2. Related Standards Activity

• M1.4 Task Group for Biometric Performance Testing and Reporting

• M1.1 Task Group for Data Formats

5. Units of Measurement Used in the Standard

The units of measurement used in the Standard will be determined during its development.

6. APPENDIX: BioAPI Version 1.1 Section 2.1.46

A value indicating the quality of the biometric data in a BIR.

```
typedef sint8 BioAPI_QUALITY;
```

NOTE: All integer values in the BIR header are little-endian.

The performance of biometrics varies with the quality of the biometric data. Since a universally accepted definition of quality does not exist. BioAPI has elected to provide the following structure with the goal of framing the effect of quality on usage of the BSP (as envisioned by the BSP vendor). The scores as reported by the BSP are based on the purpose (BIR_PURPOSE) indicted by the application (e.g. Capture for enrollment/verify, capture for enrollment/identify; capture for verify, etc.). Additionally, the demands upon the biometric vary based on the actual customer application and/or environment (i.e. a particular application usage may require higher quality samples than would normally be required by less demanding applications).

Quality measurements are reported as an integral value in the range 0-100 except as follows:

Value of –1: BioAPI_QUALITY was not set by the BSP (reference BSP vendor's documentation for explanation).

Value of –2: BioAPI_QUALITY is not supported by the BSP.

There are two objectives in providing BioAPI_QUALITY feedback to the application:

1. The primary objective is to have the BSP inform the application how suitable the biometric sample is for the purpose (BioAPI_PURPOSE) specified by the application (as framed by the BSP vendor based on the use scenario intended by the BSP vendor).

2. The secondary objective is to provide the application with relative results (e.g. current sample is better/worse than previous sample).

Quality scores in the range 0-100 have the following interpretation:

0-25: UNACCEPTABLE: The biometric data cannot be used for the purpose specified by the application (BioAPI_PURPOSE). The biometric data must be replaced with a new sample.

26-50: MARGINAL: The biometric data will provide poor performance for the purpose specified by the application (BioAPI_PURPOSE) and in most application environments will compromise the intent of the application. The biometric data should be replaced with a new sample.

51-75: ADEQUATE: The biometric data will provide good performance in most application environments based on the purpose specified by the application (BioAPI_PURPOSE). The application should attempt to obtain higher quality data if the application developer anticipates demanding usage.

76-100: EXCELLENT: The biometric data will provide good performance for the purpose specified by the application (BioAPI_BIR_PURPOSE). The application may want to attempt to obtain better samples if the sample quality (BioAPI_QUALITY) is in the lower portion of the range (e.g. 76, 77,…) when convenient (e.g. during enrollment).