INCITS/M1-03-0032

InterNational Committee for Information Technology Standards
INCITS Secretariat, Information Technology Industry Council (ITI)
1250 Eye St. NW, Suite 200, Washington, DC 20005
Telephone 202-737-8888; Fax 202-63-4922
email: ncits@itic.org

Title: **The NIST HumanID Evaluation Framework**

Source: **National Institute for Standards and Technology (NIST)**

Submitted by: **R.M. McCabe**

Date: **31 January 2003**

Version: **1.0**

Authors: **Ross J. Micheals, Patrick Grother, and P. Jonathon Philipps**

Image Group, Information Access Division, NIST

rossm@nist.gov

Purpose:

This document describes a method and evaluation process for the quantitative testing of biometric recognition systems. This framework was used to facilitate the administration of the Face Recognition Vendor Test 2002 (FRVT2002). Document M1-03-0025, Performance Metrics for FRVT2002 presents the general framework used by NIST researchers for the quantitative determination of the performance metrics for biometric recognition systems. Both of these documents are contributions to the Ad-Hoc Group on Performance Testing and Reporting (AHGPTR) and are being submitted for consideration as part of the USNB's position.

The NIST HumanID Evaluation Framework

Ross J. Micheals, Patrick Grother, and P. Jonathon Phillips

Image Group, Information Access Division, National Institute of Standards and Technology, Gaithersburg MD 20899, USA,
rossm@nist.gov

Abstract. The NIST HumanID Evaluation Framework, or HEF, is an effort to design, implement, and deploy standards for the robust and complete documentation of the biometric system evaluation process. The HEF is an attempt to leverage contemporary technologies, specifically XML, for the formal description of such tests. The HEF was used to facilitate the administration of the 2002 Face Recognition Vendor Test, or FRVT 2002. Unlike FRVT 2000 or FERET 96, FRVT 2002 used both still and video facial imagery, warranting the development of a more sophisticated and regular means of describing data presented to the participants.

1 Introduction

The HumanID Evaluation Framework, or HEF, is a mechanism for the quantitative testing of biometric recognition systems. It is an extension of the FERET protocol, which defined a framework for the evaluation of face recognition technologies. The HEF, however, is general enough to apply to virtually any recognition task, and can be applied to arbitrary, heterogeneous mixtures of biometric systems.

The goal of the HEF effort is to design, implement, and deploy standards for the robust and complete documentation of the biometric system evaluation process. A mandate of the scientific method is experimental repeatability — achievement of this goal in the evaluation of biometric systems is often thwarted by inadequate documentation of experimental procedures.

The initial version of HEF, presented to the public for the first time here, was coupled with the 2002 Face Recognition Vendor Test, or FRVT 2002. Although FRVT 2002 has played a key role in shaping the current form of the HEF, it should be noted that they are separate, but related efforts. FRVT 2002 provided a unique opportunity for the development and implementation of an "initial" version of HEF. NIST has already applied the HEF for an internal evaluation of a fingerprint matcher.

The primary focus of HEF is the documentation of the input test suites and output recognition hypotheses, and *not* the algorithms embedded within a particular recognition system. In various domains (character recognition, fingerprint matching, automatic target recognition), much effort has been put into the gathering of standard training and testing suites that facilitate algorithm

development and subsequent evaluation. The use of test data for empirical testing is widespread in the scientific literature. The HEF is an attempt to leverage contemporary technologies for the formal description of such tests. Accordingly, the HEF defines a suite of XML-based markups for the inputs to, and outputs of, recognition systems. Unlike the XML Common Biometric Format, or XCBF [1], the focus of the HEF is the *evaluation* of biometric systems, and not the biometric information itself.

The HEF is designed to facilitate off-line, black-box empirical testing. A recognition engine takes two sets of biometric signatures, the enrolled and the "unknown" test samples, and produces some form of identification data. Currently, the HEF assumes that this output data is a collection of scores, with each score indicating the similarity between a pair of signatures.

The main goal of this initial version of the HEF is to have a well-defined means of marking up sets of biometric data. This paper describes the markup of "signature sets," which list the biometric data from a group of individuals. For added flexibility, HEF provides not a single schema, but a *family* of schemas that may be used to validate different kinds of signature sets. Using a set of schemas allows applications to validate signature sets using domain-specific criteria, while still maintaining a high-level, and global, consistency.

The original description of the HEF included an XML format that may be used to describe the raw signature-to-signature similarity values. However, in the FRVT 2002 High Computation test alone, participants were required to provide over 1.36 *billion* similarity values. Even when the participants saved the results in binary form, with 4-bytes per floating-point result, and minimal metadata, over 60 GB of data per vendor was produced. Requiring that the participants output their similarity scores in XML was not maintained as a viable option.

The main goal of this paper is to introduce to the international biometric community, the details of the XML protocol used for formally describing the biometric data used for FRVT 2002. It is not meant to be a complete description of the protocol, rather, an introduction to the abstractions required for a generic description of biometric evaluation data.

2 Terminology

In the HEF model, the following terminology describes biometric information at different grouping levels. Each human subject of interest is an *individual*. A collection of biometric data for a single individual makes up a *signature*. A collection of signatures constitutes a *signature-set*.

For a given individual, a particular biometric recording *event* corresponds to a *sigmember*. An event in this context is typically a time-localized period during which the subject is imaged. Since an individual may have many biometric recordings, a single *signature* can contain one or more *sigmembers*. For example, in the FERET database, there are images of some subjects taken several months apart. In this case, each image is a different *sigmember*. In the general case a

signature will be comprised of *sigmembers* that contain heterogenous biometric data; for example, a fingerprint and a mugshot.

A *sigmember*, or recording event, could also contain one or more data components. For example, a stereoscopic video might consist of two (simultaneously captured) video sequences. A *dataset* corresponds to a logical component of a biometric recording. It follows that a *sigmember* may contain one or more *dataset*s. The precise definition of a dataset is expected to change according to the mode of biometric. For most biometrics, however, a single dataset is often sufficient.

Under the HEF, it is assumed that for each *dataset*, there exists a set of one or more files containing the raw biometric data of interest. Therefore, each *dataset* may contain one or more *file*s. Each *file* corresponds to a data file that contains biometric data. Note that the HEF does not attempt to restrict the permissible file formats (JPEG, PNG, MPEG, AVI, etc.) in any way.

In summary, the sequence of terms (*signature-set*, *signature*, *sigmember*, *dataset*, *file*), in order from most abstract to most concrete, define a heirarchy designed to separate the test structure from the underlying datafiles

3 Detailed Example

To illustrate the above terminology, consider the following example. Suppose we have biometric data on three different subjects — Patrick, Ross, and Jonathon – and we wished to create a structured document that describes these signatures.

- **Patrick**. Suppose Patrick's data consists of is a **single facial image**. Then Patrick's signature a single sigmember, with a single dataset, with a single file that contains an image of Patrick's face.
- **Ross**. Suppose Ross' data consists of a **short video clip**. The video, however, is not stored in a single file, but as a collection of five individual frames or images. Then, Ross' signature has a single sigmember, with a single dataset, with five files that each contain a different frame. For this subject, there is only one sigmember since the video clip is from a single recording event. There is also only one dataset, since the individual frames, are part of the a larger logical component — the "video."
- **Jonathon**. Suppose Jonathon's biometric data includes an **iris scan**, **three facial images** each taken on different days, and a **stereoscopic gait video**. Jonathon's signature therefore contains five sigmembers: one for the iris scan, three for each facial image, and one for the gait video. For the first sigmember, the iris scan, there is a single dataset with a single file that contains the iris data. Three sigmembers, for the facial imagery, each have a single dataset, each with a single file that each contain a facial image. The fifth sigmember, the gait video, has *two* datasets — one for each video. The datasets would each have a single file if the data was encoded in a single video (such as an MPEG), or could have a collection of files, where each file corresponded to a particular frame.

In the next section, we describe an example document marking up the biometric information for these three synthetic signatures. After the overview, we give a detailed treatment of each element and attribute.

3.1 Document Overview

Figure 1 shows an example of how one would mark up the above information using the HEF signature set schemas. The non-whitespace lines have been numbered so that we can easily reference the different parts of the document — a real signature set document would not contain these line numbers. The first line of the document is a standard XML header.

Lines 2 through 6 and line 59 compose the opening and closing root elements of the document. The <*signature-set*> element also contains several attributes used for XML namespace "bookkeeping". We will briefly describe them, but for most signature sets, copying these lines verbatim will most likely be sufficient — it is not necessary to completely understand them. On line 2, the *xmlns* attribute defines the "target" namespace of the document. This associates the elements of the document with the string, or "namespace" *http://www.nist.gov/humanid/hef/xml/0.99.0*. On line 3, the *xmlns:xsi* attribute associates the prefix *xsi* with a standard name understood by XML parsers which allows the prefix *xsi* to be used to access XML Schema elements and attributes. Lines 4 and 5 indicate the location of the signature set schema. It is a pair of strings consisting of a namespace (which matches the *xmlns* value) and the (local) schema filename. Line 6 is a useable name for this particular colleciton.

Lines 7 through 13 describe the biometric information for Patrick, the first subject. This signature contains a single sigmember (lines 8 and 12), a single dataset (lines 9 and 11), and a single file (line 10). As indicated by lines 8 and 9, the data is a JPEG digital still.

Lines 14 through 25 describe the biometric information for the second individual. Ross' signature consists of a single video recording, so their signature contains a single sigmember (line 15, 16, and 24), and a single dataset (lines 17 and 23). Each frame of the video clip, however, is stored in its own file, and therefore, the dataset (line 17) contains multiple file elements (lines 18–22).

Lines 26 through 58 describe the biometric information for Jonathon, our third subject. Recall that for Jonathon we have biometric information collected via five separate recordings, and therefore, there are five separate sigmembers. Lines 27–31 are a sigmember for the iris recording information. It contains a single dataset (lines 28 and 30) and a single file (line 29). Lines 32 through 47 mark up the next three sigmembers. Since they are face stills, they are similar in structure to Patrick's sigmember (lines 8–12). We reiterate that because they are stills taken at different times — i.e., they are different recordings — each face still is a different sigmember (as opposed to using different <*dataset*> or <*file*> tags). Finally, lines 48–57 describe the stereoscopic gait video. There is only one sigmember (lines 48–57) since both videos were taken during a single biometric recording event. Each video, however, is a logical component of this recording, and therefore each corresponds to a dataset element (lines 49–52 and

lines 53–56). The files themselves, one for each video, are marked up in lines 50–51 and lines 54–55.

4 Document Structure

Because of the wide variety of biometric systems, and the varying nature of the constraints that a recognition system may want to apply on a set of signatures, the HEF includes a family of related schemas that can be used "as is" for face recognition systems, or easily extended to accommodate new ones. HEF uses *derived types* [2] as its main vehicle for accomplishing this kind of flexibility. The structure of an HEF signature set document follows directly from the above grouping terminology. Each term corresponds to a tag, and the containment relationship is represented by the nesting of elements. For example, since a signature set contains multiple signatures, the <*signature-set*> may have one or more <*signature*> elements as child elements — a <*signature*> element may have one or more <*sigmember*> elements as children — and so on. In the remainder of this section we detail the role of each element in a signature set document.

4.1 Signature Set & Signature Elements

The root tag of every valid signature set document must be <*signature-set*>. The optional *name* attribute is a token that may be used to name the signature set, so that an application can refer to a particular signature set by using its name, as opposed to its filename, or other characteristic. A valid <*signature-set*> must have one or more valid child <*signature*> elements. If a <*signature-set*> does not have at least one <*signature*> child element, then the document will not validate successfully.

Each <*signature*> corresponds to a collection of biometric information for a single individual. Each signature must have a unique *name* which is a token that can be used to refer to a particular signature. An optional attribute *subject_id* is a token associated with a particular *subject*, as opposed to a *signature*. It is important to understand the difference between *name* and *subject_id*. Given the nature of an evaluation, the *name* may or may not contain information about the subjects true identity. In an external evaluation, there may be a need to hide a subject's identity from a recognition system, and it would be expected that the *subject_id* attribute not be provided. However, there is still a need to provide handles for specific signatures — this is what the *name* attribute is for. For FRVT 2002, the signature *name*s were obfuscated to prevent participants from correctly identifying subjects based on the XML document alone.

To ensure that both signatures and individuals can be referred to without ambiguity, within a single signature set file, no pair of <*signature*> elements may share the same *name* value. A valid <*signature*> contains one or more valid <*sigmember*> elements. The enforcement of uniqueness constraints across different files is an option of an application.

```
 1:  <?xml version="1.0" encoding="UTF-8"?>

 2:  <signature-set xmlns="http://www.nist.gov/humanid/hef/xml/0.99.0"
 3:                 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 4:                 xsi:schemaLocation="http://www.nist.gov/humanid/hef/xml/0.99.0
 5:                                     0.99.0/sigset-schemas/sigset-unrestricbted.xsd"
 6:                 name="example set">

 7:   <signature name="Patrick's signature" subject_id="Patrick">
 8:    <sigmember xsi:type="simple-face-still-type" modality="simple face still">
 9:     <dataset media="digital still" type="jpeg">
10:      <file name="patrick-01.jpg"/>
11:     </dataset>
12:    </sigmember>
13:   </signature>

14:   <signature name="Ross' signature" subject_id="Ross">
15:    <sigmember xsi:type="simple-multifile-face-video-type"
16:               modality="simple multifile face video">
17:     <dataset>
18:      <file name="ross-seq01-frame00.jpg"/>
19:      <file name="ross-seq01-frame01.jpg"/>
20:      <file name="ross-seq01-frame02.jpg"/>
21:      <file name="ross-seq01-frame03.jpg"/>
22:      <file name="ross-seq01-frame04.jpg"/>
23:     </dataset>
24:    </sigmember>
25:   </signature>

26:   <signature name="Jonathan's signature" subject_id="Jonathon">
27:    <sigmember modality="iris scan">
28:     <dataset media="digital still">
29:      <file name="jonathan-iris.dat"/>
30:     </dataset>
31:    </sigmember>
32:    <sigmember xsi:type="simple-face-still-type" modality="simple face still">
33:     <dataset media="digital still" type="jpeg">
34:      <file name="jonathon-still-01.jpg"/>
35:     </dataset>
36:    </sigmember>
37:    <sigmember xsi:type="simple-face-still-type" modality="simple face still">
38:     <dataset media="digital still" type="jpeg">
39:      <file name="jonathon-still-02.jpg"/>
40:     </dataset>
41:    </sigmember>
42:    <sigmember xsi:type="simple-face-still-type" modality="simple face still">
43:     <dataset media="digital still" type="jpeg">
44:      <file xsi:type="spatial-file-type" name="0003-03.jpg"
45:            roi="(10,30) (125,110)"/>
46:     </dataset>
47:    </sigmember>
48:    <sigmember modality="stereo gait video">
49:     <dataset media="digital video" type="mpeg">
50:      <file xsi:type="temporal-file-type" name="jonathon-gait-01-left.mpeg"
51:            start="10" stop="230" unit="frame">
52:     </dataset>
53:     <dataset media="digital video" type="mpeg">
54:      <file xsi:type="temporal-file-type" name="jonathon-gait-01-right.mpeg"
55:            start="10" stop="230" unit="frame">
56:     </dataset>
57:    </sigmember>
58:   </signature>

59:  </signature-set>
```

Fig. 1. The signatures for the authors described in valid HEF XML.

4.2 Sigmember Elements

Each <*sigmember*> corresponds to a particular biometric recording event — the process of collecting new biometric information about a subject. There are two kinds of <*sigmember*> types. First, a <*sigmember*> element may be of the base type *sigmember-type* — a generic type that can be used to describe a recording of arbitrary mode. The other types are defined as restricted derivations of *sigmember-type*. Currently, there are three such types, *simple-face-still-type*, *simple-face-video-type* and *simple-multifile-face-video-type*, which are designed to be the preferred types for describing some common modes.

- *simple-face-still-type* for facial images stored in a single data file (e.g., digital stills or scanned photographs).
- *simple-face-video-type* for facial video stored in a single data file (e.g., an MPEG, AVI, or QuickTime video clip).
- *simple-multifile-face-video-type* for facial video stored in multiple data files (e.g., a sequence of GIFs, TIFFs, JPEGs, or PNMs).

All of these types are accessed through the use of the <*sigmember*> tag, but require the use of the special attribute *xsi:type* to indicate, to the XML parser, that they are not of the base type <*sigmember-type*>, but of a specified derived type. Typically, these derived sigmember types versions of the base type *sigmember-type*, but with additional constraints. If the derived types are insufficient for describing new biometric data, either the base type *sigmember-type* can be used, or a new derived type could be written.

The required *modality* attribute describes the modality of the recording. The *xsi:type* attribute, indicates to the XML *parser*, the proper derived type to use when validating the document. All <*sigmember*> elements may also use the optional *metadata* attribute. This attribute is a token that an application can use to reference recording meta-information, such as sensor information, persons involved in the data collection, contact information, and so on.

4.3 Dataset & File Elements

Every valid <*sigmember*> must contain at least one <*dataset*>, where each <*dataset*> corresponds to some logical component of a biometric recording. Naturally, all <*dataset*>s with a common parent should correspond to the same biometric recording. The precise definition of dataset is expected to change according to the mode of the biometric. For most biometrics, such as a facial image still, it is expected that a <*sigmember*> has a single <*dataset*>. It is anticipated that more complex recordings would contain more than one <*dataset*> (for example, one dataset per camera in the case of stereoscopic video). Until more derived sigmembers are added to HEF, it is expected that HEF users will define their own conventions and schemas for determining the nature and quantities of datasets for their own biometric recordings. There are two optional attributes that may be used with the <*dataset*> element. The *media* attribute is a token which describes the media of the original recording — “digital still”,

"35mm film", or "Hi-8 video" for example. The *datatype* attribute is a token which should be used to describe the format of the recorded data — "JPEG", "MPEG", "TIFF", and so on. Like most of the optional attributes, there is no strict convention for the values of *media* and *datatype*.

Finally, there is the bottommost tag, the <*file*> element. Each <*file*> element is terminal and cannot contain child elements. Each file element is used to refer to a particular datafile, which, depending on the context of the parent tags, may contain complete or partial data of a biometric recording.

Like <*sigmember*>, the <*file*> has a base and derived types. The base type, *file-type* is designed for general use, and has a single required attribute *name* which is a token that indicates the biometric data's file name. The two derived types provide mechanisms for indicating to an application more information about data of interest. For example, there may be several faces in a digital still, or several people captured in a gait video, but only one spatiotemporal region of interest.

The *spatial-file-type* extends the base type by offering an optional *roi* attribute, which is a token that can be used to specify a region of interest within the data file. The *temporal-file-type* is designed to be the preferred type for describing recordings with temporal information. This type requires the *start* and *stop* attributes, which are integer indexes that indicate to an application the logical beginning and ending of the biometric data. The optional *unit* attribute is a token to indicate the unit of measure to associate with the start and stop indexes. For example, in a video sequence, the unit value may be "frame". A type of *temporal-file-type* may also use the *roi* attribute.

5 Conclusions

In adminstrating FRVT 2002, the HEF played a vital role in unambigously describing sets of biometric signatures. During the development of the the FRVT 2002 scoring software, formats for descrbing *identification,* verification, and *watchlist* scenarios [3], were also developed. Schemas for these formats will be made availabe for the next public release of HEF. Further augmenting the scoring suite with various XSLTs allowed the ROC, CMC and watch-list data to be transformed into scripts for use with GNUplot, R, and Splus. Requests for information about obtaining the HEF can be e-mailed directed to the authors.

References

1. Oasis XML Common Biometric Format (XCBF). *http://www.oasis-open.org/committees/xcbf/*
2. XML Schema Part 0: Primer *http://www.w3.org/TR/xmlschema-0/*
3. P. Grother, R. Micheals, and J. Philips. Face Recognition Vendor Test 2002 Performance Metrics These proceedings.