J2 Development and Validation of an Automated Biometric Handwriting Comparison System

JoAnn Buscaglia, PhD, FBI Laboratory, Counterterrorism & Forensic Science Research Unit, FBI Academy, Quantico, VA 22135; Mark Walch, GTG LLC, 1000 North Payne Street, Alexandria, VA 22314; Donald T. Gantz, PhD, George Mason University, Department of Applied and Engineering Statistics & Department of Applied Information Technology, 157 Science-Technology II, Fairfax, VA 22030; and John J. Miller, PhD, George Mason University, Department of Applied and Engineering Statistics, 157 Science-Technology II, Fairfax, VA 22030*

After attending this presentation, attendees will learn about key technologies underlying the development, implementation, and validation of an automated handwriting-derived biometric identification system, as well as the conceptual basis for the "Biometric Kernel" of user-specific characteristics that can be used to distinguish among writers and the empirical/ statistical basis for using the kernel as a handwriting-based biometric identifier.

This presentation demonstrate how the development of an automated handwriting identification system can assist document examiners with sorting through large numbers of documents from various writers and provide potential candidate writers from previous offenses. The existence of an objectively-derived writer's biometric identity supports *Daubert* chal- lenges to the premise of individuality of handwriting, and the development of this system provides a tool for future statistical studies of handwriting individuality.

Forensic document examiners routinely perform handwriting compar- isons for writer identification. The underlying premise for such identifica- tions is that each person incorporates individual features into handwriting, which can be used to distinguish it from that of other writers. During *Daubert* admissibility hearings, the validity of this foundation of individu- ality and the significance of handwriting characteristics as associative evi- dence have been challenged. In order to assess the frequency of occurrence and significance of handwriting characteristics, a collection of handwriting samples from a large number of individuals must be acquired. Further, in order to evaluate this premise, it is advantageous to compare handwriting using an automated system, which can provide quantitative data for verifi- cation of common features and process a larger database of samples than can manual examination. This presentation provides a discussion of key technologies underlying the development, implementation, and validation of an automated handwriting-derived biometric identification system, the Biometric Handwriting Comparator (BHC). It will also address the con- ceptual basis for the "Biometric Kernel" of user-specific characteristics that can be used to distinguish among writers, as well as the empirical/statistical basis for using the kernel as a handwriting-based biometric identifier. Finally, the results of blind testing of the BHC for automated writer identi- fication utilizing small amounts of text will be presented.

Handwriting and hand printing samples from approximately 500 indi- viduals were collected. Because it is the current practice in the document community, when possible, to collect writing exemplars from individuals in a single sitting, and for convenience of sampling such a large population of volunteers, exemplars for this project were collected in this manner. Exemplars were collected from volunteers at various locations, including government agencies, professional meetings, and via personal contacts. Each participant was given a packet containing a writing passage (modified London business letter), personal information form, informed consent form, ten sheets of unlined paper and a Bic® pen (medium black RoundStic™). Exemplars from each participant consist of the passage written five times in cursive (script) and five times in hand printing on sep- arate pages. Collected exemplars were optically scanned and used as ref- erence and test sets in automated BHC. These writing exemplars, which provide a measure of both intra- and inter-writer variability, aided in the development and validation of the BHC for the comparison of writers. While these samples of convenience are suitable for this purpose, addi- tional targeted sampling is needed to fully explore the premise of the indi- viduality of handwriting.

The BHC is a portable, PC-based system that permits automated, real- time identification of the writer of a document. Handwriting-derived bio- metric identification exploits the rich set of measurements available through Isomorphic Graph Matching, a technique based on Graph-Theory that is used to identify the same written forms in different writing samples. By statistically comparing measurements on similar objects across different writing, identifying those writing characteristics that best distinguish or characterize individual writers is possible. A writer's biometric identity, or "Biometric kernel," is defined through the measurements that are deter- mined to characterize that person's writing, in the sense that those mea- surements have the power to distinguish his/her writing from that of other writers. Handwriting-derived biometric identification is a computationally intense process that utilizes statistical discrimination algorithms.

The BHC with statistical discrimination algorithms has been tested uti- lizing text with large and small quantities of characters with good results. Using a database of cursive handwriting from 100 writers, when a sample of 15 or more characters was used to locate a true writer, over 95% of the time that true writer had the highest score. That is, the true writer was iden- tified by testing as the most likely writer for the sample of characters. Additionally, the true writer was virtually always identified as one of the two top scoring candidate writers.

In order to test the system under a more realistic case scenario, a blind test is currently underway using mock bank robbery notes collected from various individuals, both writers who previously submitted samples and some not currently in the database. These writing samples were collected using unlined paper and a blue or black ink

 ** Presenting Author*

pen, which was chosen by the writer. The bank robbery note exemplars were collected almost three years after the original reference documents to which they will be compared.

Automated Handwriting Identification, Forensic Document Examination, Handwriting Individuality

 ** Presenting Author*