J4 Evaluation of the Individuality of Handwriting Using FLASH ID – A Totally Automated, Language Independent System for Handwriting Identification

Mark A. Walch, MPH, The Gannon Technologies Group, 1000 North Payne Street, Alexandria, VA 22134; Donald T. Gantz, PhD*, George Mason University, Department of Applied Information Technology, MS 1G8, 4400 University Drive, Fairfax, VA 22030; John J. Miller, PhD and; Linda J. Davis, PhD, George Mason University, Department of Statistics, 4400 University Drive, MS 4A7, Fairfax, VA 22030-4444; Christopher P. Saunders, PhD, George Mason University, Document Forensics Lab, 4400 University Drive, MS, 1GB, Fairfax, VA 22030; Mark J. Lancaster, PhD, and Andrea Lamas, BS, George Mason University, 4400 University Drive, MS, 1GB, Fairfax, VA; and JoAnn Buscaglia, PhD*, Federal Bureau of Investigation Laboratory, Counterterrorism and Forensic Science Re- search Unit, Building 12, Quantico, VA 22135*

The goal of this presentation is to expose attendees to an innovative and highly effective means of automatic handwriting-derived biometric identifi- cation using the "FLASH ID" software package. Attendees will become familiar with the statistical techniques behind this software. They will learn how a writer is characterized through quantitative analysis of a writing sample and they will see how a data base of writers is scored and sorted to identify the writer of a questioned document from among the population of writers in a data base. Finally, attendees will learn about the validation of the concept of handwriting uniqueness using empirical studies of measurable features extracted from handwriting samples.

This presentation will impact the forensic community by providing information of the ongoing collaboration of three organizations who are working to use automated biometric identification of handwriting tools to assist forensic document examiners (FDEs). Developers, statisticians, FDEs, and research scientists have been working to develop and apply biometric identification tools to support forensic handwriting individualizations and to address statistical issues involving individuality in the context of handwriting.

Forensic document examiners routinely perform handwriting compar- isons for writer identification. The underlying premise for such identifica- tions is that each person incorporates distinguishing individual features into his/her handwriting. During *Daubert* admissibility hearings, the validity of this foundation of individuality has been challenged. A collaborative research effort addressing this challenge has resulted in FLASH ID, a totally automated language-independent system for handwriting identification. FLASH ID uses an innovative quantification of handwritten text and computationally intense statistical methods for discrimination among writers. This presentation will consist of three presentations focusing on: (1) devel- opment and functionality of the FLASH ID system, (2) statistical methods for biometric identification with handwriting, and (3) empirical testing to assess the individuality of handwriting and how it relates to *Daubert.*

The "FLASH ID" software package will be presented as a fully opera- tional software system that can address the immediate needs within the forensic community related to using handwriting as a biometric identifier. The presenter will illustrate how individual features, available and quantifi- able within a person's writing, can be empirically captured into a "loss less" data structure that preserves the topology and geometry of the original writing. Statistical algorithms are created to reduce the very large number of feature measurements down to a very few, called a writer's "biometric kernel," that captures those elements that link the writing to its writer. The Biometric Kernel is the statistically-derived subset of those measurements that truly captures the essence of an individual's writing. Once the Biometric Kernel is established, FLASH ID can act on any unknown sample of hand- writing and will return the nearest value in its handwriting reference database that provides the closest match to the questioned writing sample. FLASH ID represents a new approach toward using handwriting as a biometric identifier that does not attempt to replicate the actions of a forensic document exam- iner. Rather, it brings to bear the power of what computers do very well—rapid capture and processing of large quantities of data—into the hands of forensic experts.

A method of quantification of handwriting originally applied to optical character recognition has been demonstrated to provide a powerful founda- tion for biometric identification using handwriting. A recognized character in a document is associated with a mathematical graph, which is an array of curves that intersect or end in vertices. The frequency pattern of graph types observed in a document for each separate alphabetic character is a very pow- erful biometric identifier of the writer of the document. For instance, using pair-wise comparisons of 292 copies of a modified London business letter written by 100 writers (approximately 3 documents per writer), this biometric identifier correctly linked the documents for each writer. A similar exercise using only segments of the London letters showed that about 50 characters suffice for identification with high accuracy. This exercise with discordant segments demonstrated that identification accuracy was context independent. Further, additional information about the writer's profile are obtained using a minutia level biometric identification, which is based on physical feature measurements on the graph associated with a character. Based on minutia level biometric kernel-based identification alone, the true writer of a questioned document will be retrieved from a data base of known writers with high accuracy if that writer's minutia characterization is stored in the data base.

As a residual biometric that can link individuals to documents, hand- writing provides an important data source for both law enforcement and intelligence purposes. In the form of FLASH ID, the forensic science community will now have a tool that harnesses the power of automation to leverage the effectiveness of document

 ** Presenting Author*

examiners by capturing similarities embedded among multiple writing samples and graphically showcasing these similarities supported by the statistical analysis that led to their identification. The technology underlying FLASH ID is language independent; that is, the empirical and analytical techniques that power the handwriting-derived biometric process have been demonstrated to function in different languages with completely different scripts. In this way, FLASH ID will extend document forensics across language barriers—something that is not com- monly practiced today.

FLASH ID represents a totally automated process for extracting graphical data from handwritten documents, analyzing this data using robust statistical methods and matching documents based on similarity of the captured writing. This presentation will highlight the high level features of this research, supported by more detailed poster presentations regarding the FLASH ID system, statistical characterization of handwriting using features derived with FLASH ID, and statistical concepts for assessing handwriting individuality.

Handwriting, Individuality, Statistics

 ** Presenting Author*