J1 Validation of an Automated Handwriting- Derived Biometric Identification System (FLASH ID) on English and Arabic Writings

Mark A. Walch, MBA, The Gannon Technologies Group, 1000 North Payne Street, Alexandria, VA 22134*

The goal of this presentation is to illustrate an automated technique for handwriting derived biometric identification that transcends languages and scripts. The biometric power of handwriting is embedded in graphemes - compact graphical forms - that exist in all languages.

This presentation will impact the forensic science community by empowering forensic document examiners to new frontiers—by extending their reach into new languages - that have previously been perceived as barriers.

This presentation will provide updated information regarding FLASH ID; a highly effective means for automatic handwriting derived biometric identification. The "FLASH ID" software package was developed with extensive guidance and technical input provided by forensic scientists from the Federal Bureau of Investigation Laboratory.

A specific aspect of FLASH ID's functionality will be presented in the form of its "language independence." Since FLASH ID decomposes writings into individual "graphemes" which transcend individual languages, it can perform writer identification in multiple languages either concurrently or independently. This ability to extend Forensic Document Examination across multiple languages represents a significant breakthrough over current practices which tend to be language specific.

The audience will learn that the biometric power of handwriting does not rest with individual characters or words, but takes the form of parts of characters and character connectors. Handwriting biometric power also does not rely on any particular script, but rather, it crosses both languages and scripts. The mathematical underpinnings of FLASH ID are based on Graph Theory and FLASH ID works by quantifying graphical features, available within an individual's writing, into a "loss less" data structure that preserves the topology and geometry of the original writing. The methods for creating this data structure as well as the actual structure are "language agnostic." That is, the data structure is built from graphemes—graphical forms—rather than actual characters from a particular language. This data structure captures both the topology and hundreds of detailed physical measurements from written forms. Using this graph-based format, FLASH ID employs statistical methods to distill the topological and physical features into a "biometric kernel." The Biometric Kernel captures the essence of the repertoire of physical forms used by a particular individual to write in a particular language.

Specific experiments to validate the language independent capabilities of FLASH ID when applied to the identification writers in a language other than English will be discussed. For purposes of this presentation, Arabic is the chosen language. Topics to be presented include: similarities and dissimilarities in the graphical composition of language; empirical evaluations of grapheme distributions between languages; biometric writer identification performance on collections of English, Arabic, and mixed commingled documents; and the feasibility of capturing class characteristics of languages. The presentation will culminate in a live demonstration showing FLASH ID applied to both English and Arabic documents in the same collection.

A key point to be made to the audience is FLASH ID represents a new approach toward handwriting examination that will empower forensic document examiners to handle documents in multiple languages. The core message will be rooted in two important aspects of the technology used to build FLASH ID. First, FLASH ID represents a totally automated process for extracting biometric data from handwritten documents regardless the source language, analyzing these data using established statistical methods and matching documents based on similarity of the captured writing to known writing. Second, the technology underlying FLASH ID's language independence has been demonstrated to function in different languages with completely different scripts.

As a residual biometric that can link individuals to documents they have written, handwriting provides an important data source for both law enforcement and intelligence purposes. FLASH ID provides the forensic science community with a tool that harnesses the power of automation to make current practices more efficient and effective. By transcending language, the impact of FLASH ID is that it will empower forensic document examiners to new frontiers—by extending their reach into new languages—that have previously been perceived as barriers.

Handwriting, Biometric, Arabic

 ** Presenting Author*