J6 The State of the Art in Computational Forensic Document Examination

Timothy Campbell, BSc, Canada Border Services Agency, 14 Colonnade Road, Ste 280, Ottawa, ON K2E7M6, CANADA*

The goal of this presentation is to provide attendees with a better understanding of the document research work being conducted that may not initially have a forensic intent but may ultimately be applicable to further development of the profession.

This presentation will impact the forensic science community by alerting practitioners to tools and techniques they would not normally find within the forensic community.

Forensic Document Examination (FDE) is a broad discipline that, while necessarily being firmly tied to traditional technologies, is greatly affected by modern advancements in instrumentation and analytical techniques. FDEs are quite familiar with databases for examined subject matter, such as inks and typewriters, and digital tools, such as laser microscopes, spreadsheets, and advanced image enhancement; however, there is another parallel world of work conducted on documents in academia that FDEs should find beneficial.

The Pattern Recognition (PR) research community has been working on applying PR techniques to document and handwriting problems as they have been presented in the world of archiving and commerce. While these subjects tend to focus on digitizing historical tomes or modern business documents, a subset of PR researchers have worked on automating handwriting authorship determinations, the same analyses that FDEs claim as expertise.

The International Association for Pattern Recognition (IAPR) has a number of Technical Committees (TC); number 11 (TC11 – Reading Systems) concerns the analysis and recognition of document-based information, namely image processing, Optical Character Recognition (OCR), and handwriting recognition. TC11 publishes the *International Journal on Document Analysis and Recognition*. There are two recurring conferences and one workshop which are held biannually and in alternate years.

The International Conference on Document Analysis and Recognition (ICDAR) is held on odd calendar years and is geared primarily, but not exclusively, toward non-handwriting document issues. On even years, the International Conference on Frontiers in Handwriting Recognition (ICFHR) is held, as is the main TC11 workshop, Document Analysis Systems (DAS).

Smaller workshops held in conjunction with these conferences have included Historical Image Processing (HIP), the International Workshop on Camera-Based Document Analysis and Recognition (CBDAR), the International Workshop on Multilingual OCR (MOCR), and the International Workshop on Robust Reading (IWRR). New for 2017 are the Historical Book Analysis (HBA) workshop and the First International Workshop on Computational Document Forensics (IWCDF). Additionally, the Automated Forensic Handwriting Analysis (AFHA) group was well-attended by FDEs when it was co-located with the meeting of the American Society of Questioned Document Examiners (ASQDE) in 2014.

These gatherings typically make use of competitions whereby interested parties will acquire a common dataset of known ground truth and apply their algorithms against the problem at hand. Winners are announced after the submissions have been evaluated for efficiency and accuracy.

In addition, the forecast of the paperless office has largely failed to materialize and the amount of hardcopy documents used in society remains large, while we have simultaneously seen a swelling of digital document quantities. Outside of PR, the task of organizing and navigating such cross-media collections is a point of important research as institutions see the reduction of librarians as a cost-cutting measure, while the work is downloaded to the subject matter expert. Researchers are at work integrating large collections and speeding the re-finding of hardcopy documents through searching a digital collection. Moreover, as cases grow in size and complexity, such interfaces could aid the FDE in an examination capacity so that comparing handwriting and other features across thousands of documents becomes a more attainable feat.

These products and techniques may prove useful to FDEs in future casework situations or in the organization of vast reference materials and databases. A review of the state-of-the-art in computational document forensics is presented, in addition to a recap of known technologies, highlighting capabilities and limitations. The potential uses of such advanced computational tools will be discussed.

Documents, Computation, Pattern Recognition