Proc. Int. Wkshp. on Computational Forensics, Manchester, UK, Aug 31, 2007.t

Computational Forensics: Towards Hybrid-Intelligent Crime Investigation

Katrin Franke
Norwegian Information Security Laboratory,
Gjøvik University College, Norway.
kyfranke@ieee.org

Sargur N. Srihari
CEDAR, University at Buffalo,
State University of New York, USA.
srihari@cedar.buffalo.edu

Abstract

In recent years, mathematical, statistical and computer science methods have found extensive application in developing new procedures for crime investigation, prosecution and the enforcement of law. Computer-based methods have also become important tools for performing certain forensic functions. Yet, knowledge and intuition of the human expert plays a central role. This paper gives a brief overview of the potential of computational methods in forensics. The state of the art is described, some practical examples are given and a viewpoint is expressed that the methods can be extended to functions that currently rely only on expert knowledge.

1. Introduction

There are many reasons for studying the application of computational methods in the broad area of forensic sciences. The most important one is that computational methods provide a powerful toolset for representing human expert knowledge and for implementing recognition and reasoning abilities in machines. They allow for tools that help overcome the limitation of human capabilities. The concepts of machine learning and computational intelligence play a key role. Another reason for studying computational methods is that courtroom forensic testimony is often criticized by defense lawyers as lacking a scientific basis [11]. Computational methods enable the forensic practitioner to:

– analyze and identify traces in an objective and reproducible manner,
– assess the quality of an examination method,
– report and standardize investigative procedures,
– search large volumes of data efficiently,
– visualize and document the results of analysis,
– assist in the interpretation of results and their argumentation,
– reveal previously unknown patterns / links, to derive new rules and contribute to the generation of new knowledge.

Computational methods are not meant to replace the human forensic examiner, at least in the foreseeable future! They are intended to (i) assist in basic and applied research, e.g. to establish or prove the scientific basis of a particular investigative procedure, and (ii) support the forensic examiner in his / her daily casework. Modern crime investigation shall profit from the hybrid-intelligence of humans and machines.

The objective of this paper is to lay the foundations and to encourage further discussions on the development of computational methods for forensic investigation services. Researchers and practitioners in computer science are introduced to specialized areas and procedures applied in forensic casework. Current forensic challenges that demand the development of next-generation equipment and tools are exposed. The forensic scientist and practitioner, on the other hand, are provided with an overview of fundamental techniques available in the computing sciences. Selected examples of successfully implemented computing approaches will help to gain trust in methods and technologies unknown thus far. These examples may also inspire / reveal further forensic areas that can be supported by computer systems.

The remainder of this paper is structured as follows: forensic sciences are briefly described in Section 2. In Section 3 the relevant areas of computational / machine intelligence are summarized. Section 4 aims to provide a definition of computational forensics. Some practical and ongoing research examples are provided in Section 5. Section 6 concludes with discussions and points to further directions.

2. Forensics

Forensic methods consist of multi-disciplinary approaches to perform the following tasks: (i) investigate and to reconstruct a crime scene or a scene of an accident, (ii) collect and analyze trace evidence found, (iii) identfy / classify / quantify / individualize persons, objects, processes, (iv) establish linkages / associations and reconstructions, and (v) use those findings in the prosecution or the defense in a court of law. While forensics has mostly dealt with previously committed crime, greater focus is now being placed on analyzing data gathered to prevent future crime and terrorism [10].

Forensic methods are widespread in the scientific disciplines: biology, chemistry, physics and medicine, and more specialized pathology, anthropology, ballistics to mention a few. With the advancement of criminal activities, further disciplines are getting involved as for example computer science, engineering and economics. One proposal for categorizing them is given by Saks [11], who distinguishes the classical forensic identification sciences based on individualisation (to identify a finger, a writer, a weapon, a shoe that left the mark) and the more practical-oriented disciplines based on classification and quantization (chemical, biological, medical, or physical methods) like forensic toxicology.

Forensic experts study a broad area of objects, substances (blood, body fluids, drugs), chemicals (paint, fibers, explosives, toxins), tissue traces (hair, skin), impression evidence (shoe or finger print, tool or bite marks), electronic data and devices (network traffic, e-mail, images). Some further objects to be studied are fire debris, vehicles, questioned documents, physiological and behavioral patterns.

The work process of an examination can be summarized as follows: Crime-scene investigation (CSI); Documentation / photographing of the scene; Questioning witnesses; Identification / collection and preservation of evidence; Analysis of evidence (e.g. in the laboratory); Data integration; Link analysis; Crime-scene reconstruction; Report writing and presentation of findings in court.

Forensic sciences are challenged by the fact that only *tiny pieces of evidence* are hidden in a mostly *chaotic environment*. Examples are a smudged fingerprint on a glass, a half ear print on a door, a disguised handwriting or an unobtrusive paint scratch. The majority of criminals invest all their knowledge and expertise to cover their activities and potential results. Traces have to be studied to reveal *specific properties* that allow for example to identify a person or to link a tool to a caused damage. Moreover, traces found will be *never identical* to known specimen in a reference base, even if traces are caused by the identical source. For example, producing exactly the same tool mark is impossible and printing exactly the same document is impossible. As a consequence, reasoning and deduction have to be performed on the basis of *partial knowledge, approximations, uncertainties* and *conjectures*.

In addition to human forensic expertise, the investigative procedure and employed technology decide case resolution. A forensic expert compares traces of evidence on the basis of well-defined sets of characteristics that are primarily based upon domain knowledge and personal experience. Despite great efforts to provide adequate expert training, some forensic methodologies have frequently been criticized, in particular the lack of studies on validity and reliability [11, 14]. Attempts have been made to support traditional methods with semi-automatic and interactive systems on the basis of measurements and decisions that lack objectivity and verifiability. Although promising research has been done, computer-based trace analysis is rarely applied in daily forensic casework. Rare exceptions are the fields of digital / computer forensics that use computational methods intrinsically, DNA analysis that takes advantage of algorithms originating from bioinformatics, and databases (e.g. for paint or fine arts), which use mainly manually entered meta information (verbatim) and keywords for data retrieval instead of realistic object presentations and from that derived machine-processable characteristics. Thus necessitating a study of whether forensic sciences can benefit from recent technological developments.

3. Computational / Machine Intelligence

Forensic methods can be assisted by algorithms and software from several areas in the computer science. Some of these are:

- Signal / Image Processing: where one-dimensional signals and two-dimensional images are tranformed for the purpose of better human or machine processing,
- Computer Vision: where images are automatically recognized to identify objects,
- Computer Graphics / Data Visualization: where two-dimensional images or three-dimensional scenes are synthesized from multi-dimensional data for better human understanding,
- Statistical Pattern Recognition: where abstract measurements are classified as belonging to one or more classes, e.g., whether a sample belongs to a known class and with what probability,
- Data Mining: where large volumes of data are processed to discover nuggets of information, e.g., presence of associations, number of clusters, outliers in a cluster,
- Robotics: where human movements are replicated by a machine, and
- Machine Learning: where a mathematical model is learnt from examples.

Much of computational / machine intelligence is dominated by statistically based algorithms. These methods are ideally suited to forensics where there is a need to demonstrate error rates and calculate probabilities [1].

4. Computational Forensics

Computational Forensics (CF) is an emerging interdisciplinary research domain. It is understood as the hypothesis-driven investigation of a specific forensic problem using computers, with the primary goal of discovery and advancement of forensic knowledge. CF works towards (1) in-depth understanding of a forensic discipline, (2) evaluation of a particular scientific method basis and (3) systematic approach to forensic sciences by applying techniques of computer science, applied mathematics and statistics. In practice it involves modeling and computer simulation (synthesis) and / or computer-based analysis and recognition in studying and solving forensic problems.

Several terms are currently used to denote mathematical and computing approaches in forensics. *Forensic Statistics* and *Forensic Information Technology* have the longest tradition, yet they are specific. The terms *Forensic Intelligence* and *CF* cover a broader spectrum. It is necessary to establish a sound conceptual framework for CF as in the case of computational vision, computational science, computational medicine, computational biology, etc. The term CF is preferred as it indicates formalization of the methods used by humans, analogous to the use of the term computational vision used by researchers trying to understand biological vision [9].

A systematic approach to CF ensures a comprehensive research / development / investigation process that remains focused on the needs of the forensic problem. The process typically includes the following phases: Analysis of the forensic problem and identifying the goals of study (alternate hypotheses); Determination of required / given preconditions and data; Data collection and / or generation; Design of experiments; Study / selection of existing computational methods and / or adaptation / design of new algorithms on demand; Implementation of the experiment including machine learning and training procedures with known data samples; Evaluation of the experiment as well as test of the hypotheses.

The expected impact of CF is potentially far reaching. Most obvious contributions are to:

– increase *efficiency* and *effectiveness* in risk analysis, crime prevention, investigation, prosecution and the enforcement of law, and to support standardized *reporting* on investigation results and deductions.

– perform *testing* that is often very time consuming. By means of systematic empirical testing scientific foundations can be established. Theories can be implemented and become testable on a larger scale of data. Subsequently, method can be analyzed regarding their strengths / weaknesses and a potential *error rate* can be determined.

– gather, manage and extrapolate data, and to synthesize new *data sets* on demand. In forensics, unequally distributed data sets exist; there are many correct but only a few counterfeit samples. Computer models can help to synthesize data and even simulate meaningful influences / variations.

– establish and to implement *standards* for work procedures and to journal processes (semi)-automatically. Technical equipment also supports the establishment / maintaining of conceptional frameworks and terminologies used. In consequence, data exchange and the interoperability of systems become feasible.

As a new scientific discipline, approaches and studies in CF need to be peer-reviewed and published for the purpose of discussion, consequent general acceptance, and rejection by the scientific community. Scientific expertise from forensics as well as computing have to be incorporated. Methods and studies have to be reviewed for their forensic and technological correctness.

CF requires joint efforts by forensic and computational scientists with benefits to both. Regarding sharing of knowledge among forensic and computer experts, while there may be good reasons for protecting forensic expertise within a closed community, it would conflict with Daubert and other legal rulings [3], that require the investigative method as being generally accepted, having a scientific basis, etc. The relatively small community of forensic experts can hardly foster scientific bases for their methods independently. As has been successful in the traditional forensic domains (e.g. medicine, biology and chemistry) close cooperation between forensic scientists and computational scientists are possible. In the computational sciences, successful collaborations between computer scientists and biologists, chemists and linguists are known. With these precedents, forensics can benefit from knowledge, techniques and research findings in applied mathematics and computer science. Moreover, several forensic fields cover similar work procedures and tackle similar problems although their investigation objects are different. By means of shared knowledge, sophisticated computational methods can be efficiently adapted to a new problem domain. In addition, research and development in the computing sciences can profit from problem definitions and work procedures applied in forensics. For example, skilled forgeries and partial, noisy data that pose problems regarding the robustness of an automatic system. Computer scientists can gain new insights in analysis procedures while taking the perspective of a forensic expert who has expertise in his / her field of specialization. Subsequently, computational approaches undergo fine tuning to achieve superiority.

5. Application Examples

Mathematical, statistical and computer-based methods have been used before in forensics. Computer forensics (also called digital forensics) and DNA analysis are one example. Contributions to the *scientific methododological base* of handwriting and signature analysis are reported in [4, 13], while *search algorithm* are proposed in [2]. For the *synthesis of data* samples not only software methods, but also robots are used [6]. Research on the computer-based analysis of *striation patterns* that are subject of ballistic / toolmark investigations [7] is reported. *Friction ridge analysis* is probably the area that has most benefited from computational methods, with the development of automated fingerprint identification systems [8]. However much more needs to be done in the analysis of latent prints. Computer-assisted and fully automatic computer-based *link analysis and visualization* is increasingly used by banks and insurances in examining credit-card fraud and money laundry [10]. Crime-scene reconstruction using computer graphics referred to in [15]. A *conceptual framework on terminology used* by questioned document examiners is proposed [5] that was also implemented into a *reporting system* [12]. Assistance software for argumentation is discussed in [16]. The need for professionals with the abilities to develop and to apply latest computational methods demands *education and training* of current and next generation experts [15].

6. Conclusions and Future Directions

The use of computing tools in the forensic disciplines is sometimes minimal. Many improvements in forensics can be expected if recent findings in applied mathematics, statistics and computer sciences are implemented in computer-based systems. The objectives of this paper were to: (i) increase awareness of the impact of computer tools in crime prevention, investigation and prosecution – on the one hand, among forensic scientists with expertise in biology, chemistry and medicine but with limited exposure to computational science, and on the other hand, among computer scientists unaware of a challenging application domain, (ii) introduce computer scientists to the needs, procedures and techniques of forensics, and (iii) motivate studies on computational tools in forensics and encourage joint development by forensic and computer scientists.

Several support methods are needed for CF development: international forums (e.g. conference, scientific press media) to review and exchange research results, education and training to prepare current and future researchers and practitioners, and financial support for research and development.

Computational forensics holds the potential to greatly benefit all of the forensic sciences. For the computer scientist it poses a new frontier where new problems and challenges are to be faced. The potential benefits to society, meaningful inter-disciplinary research, and challenging problems should attract high quality students and researchers to the field.

7. Acknowldegement

S. N. Srihari acknowledges partial support of the Department of Justice Grant NIJ-2005-DD-BX-K012. Views expressed are those of the authors and not the opinion of the DOJ.

References

[1] C. Aitken and F. Taroni. *Statistics and the Evaluation of Evidence for Forensic Scientists*. Wiley, 2nd edition, 2005.

[2] M. Bulacu and L. Schomaker. Text-independent writer identification and verification using textural and allographic features. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 29(4):701–717, 2007.

[3] K. Foster and P. Huber. *Judging Science*. MIT Press, 1999.

[4] K. Franke. *The Influence of Physical and Biomechanical Processes on the Ink Trace - Methodological foundations for the forensic analysis of signatures*. PhD thesis, Art. Intell. Institute, Uni. Groningen, The Netherlands, 2005.

[5] K. Franke, I. Guyon, L. Schomaker, and L. Vuurpijl. WandaML - A markup language for digital document annotation. In *Proc. 9th International Workshop on Frontiers in Handwriting Recognition (IWFHR)*, Tokyo, Japan, 2004.

[6] K. Franke and L. Schomaker. Robotic writing trace synthesis and its application in the study of signature line quality. *J. Forensic Document Examination*, 16(3):119–146, 2004.

[7] M. Heizmann and F. León. Model-based analysis of striation patterns in forensic science. In S. Bramble, E. Carapezza, and L. Rudin, editors, *Enabling Technologies for Law Enforcement and Security, Proceedings of SPIE, Vol. 4232*, pages 533–544, 2001.

[8] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer, 2003.

[9] D. Marr. *Vision*. Freeman, 1982.

[10] J. Mena. *Investigative Data Mining for Security and Criminal Detection*. Butterworth-Heinemann, 2003.

[11] M. Saks and J. Koehler. The coming paradigm shift in forensic identification science. *Science*, 309:892–895, 2005.

[12] K. Schönherr. Konzeption und Prototyp eines Ausgabe- und Reportgenerators für XML-Daten aus dem Handschriftenerkennungssystem WANDA. Master's thesis, Berlin College of Technology and Business Studies, Berlin, Germany, 2004.

[13] S. Srihari, S. Cha, H. Arora, and S. Lee. Individuality of handwriting. *J. Forensic Sciences*, 47(4):1–17, 2002.

[14] Starzecpyzel. *United states vs. Starzecpyzel. 880 F. Supp. 1027 (S.D.N.Y)*, 1995.

[15] C. Veenman and M. Worring. Forensic intelligence. *Informatie*, pages 60–65, April 2007.

[16] B. Verheij. *Virtual Arguments. On the Design of Argument Assistants for Lawyers and Other Arguers*. T.M.C. Asser Press, The Hague, The Netherlands, 2005.