Extracting Evidence from Multimedia Big Data

Geradts, Z.

Publication date
2015
Document Version
Final published version

Citation for published version (APA):
Geradts, Z. (2015). *Extracting Evidence from Multimedia Big Data*. (Inaugural lecture; No. 554). Universiteit van Amsterdam.

Extracting Evidence from Multimedia Big Data

Extracting Evidence from Multimedia Big Data

Rede

uitgesproken bij de aanvaarding van het ambt van
bijzonder hoogleraar Forensic Data Science
aan de Faculteit der Natuurwetenschappen, Wiskunde en Informatica
van de Universiteit van Amsterdam
op donderdag 12 november 2015

door

Zeno Geradts

This is inaugural lecture 554, published in this series of the University of Amsterdam.

Lay-out: JAPES, Amsterdam
Photo author: Dirk Gillissen

Mevrouw de rector magnificus,
Mijnheer de decaan,
Geachte leden van het curatorium,
Geachte leden van het bestuur van de Stichting Leerstoel Criminalistiek,
Dames en heren,

I will speak further in English, since this might be the most common language that we have here today. The presentation will be available on line on http://forensic.to/oratie. Since I am currently Vice President of The American Academy of Forensic Science and also chairman of the working group Forensic IT of the European Network of Forensic Science Institutes, this presentation will focus on developments in Europe and the United States. As new chair Forensic Data Science, it is an honor to present some overview of developments and challenges we have this field. One disclaimer however, the opinions stated in this work do not necessarily represent the position of my employers or the organization were I am member of.

As a student in 1987 it was my first contact with forensic science and the Gerechtelijk Laboratorium in Rijswijk also recommended by my father who once was a lawyer, where I did some projects on replicas of bullets as well as digitizing the bullet itself for correlation with Jan Keijzer. After working at Oce in Venlo for three years on color image processing in laser copiers, I decided to go back to forensic science, and worked in several fields ranging from tool marks to firearms, as well as video image processing. It is also the place where I worked on my PhD thesis on Contents Based Retrieval from Forensic Image Database with the University of Utrecht with Max Viergever as promotor and Jurrien Bijhold as copromotor. It included research on correlation of images of toolmarks, drugs pills, shoe prints and cartridge cases of firearms. In this time there was also a strong link to Amsterdam, with Theo Gevers on the research on drug pills as well with the company Sentient Machine Research of Marten den Uyl with neural network research which was also implemented on shoeprints.

Furthermore, I worked on casework within the fields, and testified is several known cases from the first case that was the ball point case in Leiden twenty years ago where a woman was found death with a ballpoint through her eye and I testified on ballistics, to Picornie, the case van Gogh painting, as well an arson case in Schiphol and in total I investigated and reported in over five hundred cases since 1991. In this lecture I will focus on the introduction of forensic data science, forensic science within the NFI, some examples and some expectations on big multimedia data and how this might impact society.

Chair Forensic Data Science

First I would like to start with an introduction on Chair Forensic Data Science. Then I would like to give some examples how we can interpret the evidence and how to handle big data within in forensic science. I will focus also on image and video analysis.

In the last decades the amount of digital data and storage has grown very rapidly. Most citizens, the general public and criminals alike, now own devices like smart phones and computers and use them almost continuously for their profession and in their private lives. In addition CCTV cameras observe people in public spaces and private companies. As a result such devices often carry huge amounts of data on them. This data might be information from databases, multimedia information such as recorded video and audio, interaction and download logs, mobility data, location data, or links to digital information stored somewhere on the internet and linked to various social networks. This influx of digital data has a tremendous impact on the field of forensics and is leading to new research challenges.

My focus to work and liaison on methods for extracting data from digital devices during a crime (live investigation) and after a crime has been committed, and determining what kind of evidence can be retrieved from this data in relation to the crime. The main purpose is to provide forensically sound methodology and solutions to treat large collections of data as digital traces in a criminal case and use those to derive evidential strength, hypotheses, and scenarios.

I will work on methods to integrate the three disciplines of forensic multimedia analytics within the scope of the Data Science Research Center:

- store and process;
- understand and decide;
- analyze and model.

And I would like to add for forensic science the important part report and visualize for court, since it remains important that the evidence is understandable for the court.

Methods developed within the research field should lead to higher efficiency and quality of reporting of digital evidence presented to court. The scientific contribution will be related to the evidential strength of digital evidence related to the uncertain link between data and suspect. I will initiate and conduct data-intensive R&D projects by establishing national and international co-operations with innovative companies and research teams and securing the necessary funds to execute these projects. Focus will be on projects for which neither NFI nor UvA would get funding for by themselves when applying in isolation. Furthermore active on teaching within the Master of Forensic Science, track digital Forensics as well as the forensic part of Software and Network Analysis.

We can see the number of images expanding very much, and images are often used as evidence in court. Depending on the quality of images the link to the suspect can be weak or strong. We see some examples of images for biometric features of faces. However often CCTV-systems have low quality images, and suspects of, for example robberies, are aware of these systems and will shield their faces and other biometric details. In those cases also the clothes can be important to compare as well as the gait, the height and posture, however those biometric features are often weak.

Data that comes in from many sources, ranging from personal computers, backup tapes, lawful interception, smartphones, and many new devices were cameras and sensors are stored such as the drones. The development of new devices is very rapidly, so it is nearly impossible to be prepared for all new devices, so often research is needed to add the device to a knowledge database. Within the NFI we have collected information forensic traces from digital devices in digital libraries, so the solutions developed in previous cases can be used in future cases, which reduces the time of investigation.

Big data

The amount of data is growing rapidly in the world, as well as the processing power. Raymond Kurzweil, in his article "The Law of Accelerating Returns,"[1] talks about the rate of growth of computing power. Using data on past growth Ray Kurzweil extrapolated into the future. He helped to make the rate of growth real to those not used to dealing in exponentials with the following: In 2000, $1,000 could buy a computer that had the computational power of an insect. In 2010, $1,000 could buy a computer that had the computational power of a mouse. In 2020, $1,000 could buy a computer that had the computational power of a human brain. The amount of different devices connected to the internet is also thought to grow to 20 billion of devices in 3 years.

If we look into big data, the amount of data is not the main always issue. One of the issues with big data is the impossibility to process it with standard software in a timely manner. For images these issues arises very soon, since the amount of images and video grows very rapidly, and people are limited in the amount of images and hours of video they can process. Big data also means the amount of data, which can easily grow to 10 Petabytes if we would collect all digital data confiscated by the police in 2015. To imagine how much amount of data for instance 8 Terabyte is which is now often seen in practice, all of this data results in 40 km of trucks filled with paper.

According to Gartner[2], Big data is composed of the four V's:

- Volume, the data size;
- Velocity, the speed of change. The change is also visible in the rapid development of new formats and new devices where data is stored in a different file format;
- Variety, were have the different forms of data sources;
- Veracity, is the uncertainty of data. Does this data really mean what it states, is it really truthful?

However, as Gartner[3] stated in their hype cycles, the term big data is somewhat over the top of the hype, and we are now arriving in the part of disillusion, since many examples are available that it do not work always well. The issue with data mining procedures and deep learning is that relations might be found that are actually not true.

Although big data is very good at detecting correlations, correlations that[4] analysis of a smaller data set might miss, never gives if information which correlations are meaningful. A big data analysis might reveal, for instance, that from 2006 to 2011 the United States murder rate was well correlated with the market share of Internet Explorer: Both went down sharply. However, it's hard to imagine there is any causal relationship between the two.

Another example is google flu trends. After reliably providing a swift and accurate account of flu outbreaks for several winters, the theory-free, data-rich model had lost its nose for where flu was going. Google's model pointed to a severe outbreak but when the slow-and-steady data from the Centre of Disease Control arrived, they showed that Google's estimates of the spread of flu-like illnesses were overstated by almost a factor of two.[5]

There are many examples where big data solution work well[6]. Perhaps you even had the experience that your credit card company called you informing you for some real time transactions that are suspicious. Often they will block your card and send a new one, and the statements are corrected even before you have to inform the credit card company. The algorithms work with machine learning and see suspicious patterns. Visa claims is saves billions of card fraud a year by using big data algorithms.

Big data at the NFI

Also within the Netherlands Institute we use the methods for big data analysis for casework in the fields of text analysis, data profiling, financial data analysis and social network analysis. For example, with social network analysis it is possible to unmask someone using an alias or to reveal how terrorist groups work together.

Within the Netherlands Forensic Institute there is experience with a big data platform for processing evidence. In the past this was Xiraf, and it is converting now to Hansken[7]. In the past the digital evidence investigation could take months, and this has been limited to weeks, since tactical investigators can directly search in the data. Hansken should limit this to the first 48 hours were already important information can be retrieved. With future generations of Hansken the processing of large amounts of multimedia one could imagine that combining methods from for example research at the University of Amsterdam and implementing could be beneficial for the Police.

For Hansken[8] also the three implementations of security, privacy and transparency are the main cornerstones of the system. Security by design is important since one does not want to leak data, and also not that data can be modified by other parties or the system administrators. The data as imaged by the police from the devices should remain intact. Also when examining the system it is important that viruses and other malware that are examined do not infect the systems. Privacy is important since we see the amount of private data grows rapidly. Perhaps not all users realize how much information is stored on them concerning location by use of GPS and other location information in smartphones in combination with all the multimedia data that is stored on the device. Also often this is information shared with online cloud systems in often other countries.

Court

The Supreme Court of the United States[9] in 2014 “Modern cell phones, as a category, implicate privacy concerns far beyond those implicated by the search of a cigarette pack, a wallet, or a purse," said Chief Justice John Roberts. "Cell phones differ in both a quantitative and a qualitative sense from other objects that might be kept on an arrestee's person." For that reason a warrant from the court is needed for searching a phone in the United States.

On April 24th 2015 also the court of appeal in Arnhem-Leeuwarden stated that the for collecting the data of a smart phone, a court warrant is, is needed, since it otherwise violates Article 8 of the European Treaty of Human Rights.

Within the United States there is also concern on digital evidence, and Joan Sloan[10] of the University of Alabama wrote a report to congress of the United States, where he compared it with the Wild West. Though there are efforts for standardization, the validation of the results remains an issue.

An example that is given in the case Casey Anthony[11] the software gave faulty digital evidence, which was one of the main pieces of evidence in this case. In this case the software gave 84 hits on the word Chloroform, were actually it was only searched once, where this was used as evidence for a mother killing her 2 years old daughter with Chloroform.

Validation of software remains important, however as we can see is often not easy to do. Within the European Network of Forensic Institutes a new best practice manual[12] is published that also describes how to test the software and is also an effort to try to standardize and prevent.

Examples like these can also happen in the Netherlands, when people use commercial forensic software for height measurement of persons, and just trust the results of the software out of the box. We already saw that faulty evidence is possible then. Even if the software does a calculation of uncertainty of the measurement, it this should always be validated. Also the word "forensic" in a software package does not mean that developers are aware of the issues that a forensic scientist is dealing with.

The impact of faulty evidence is much worse at forensic science where someone in a criminal case is convicted for a homicide, compared to data mining for market research, where someone receives advertisements for diapers if they have no kids.

Much efforts is given at the NFI in the quality assurance for providing error-free evidence. Validation and reporting the uncertainty of the evidence are key solutions for this. Furthermore double blind investigation and checking the reports as well as using the best practice manuals is important for preventing this to happen. Also the NFI works with an externally accredited quality system. The new best practice manual for Forensic Examination of Digital Technology explicitly states that also uncertainty of the measurement within should be considered and even reports which are written in a clear and concise manner may be still be misinterpreted.

A special issue of the Journal Digital Investigation of this month by guest editors Katrin Franke and myself, will also focus on solutions and issues with big data and intelligent data analysis.

Multimedia processing

The word Multimedia has many meanings and often includes a combination of text, audio, still images, animation, location and video. All these features combined result in a data analytics. When searching through data, often the combination of images, text and location information are important to find information on when a certain image has been made and where it is has been made. Of course within forensic science one has to be aware that digital information can easily be altered, and this is sometimes a question from the court if the data has been tampered with. This is also called image forensics in literature. There exist several methods to automate this, however[13] for court cases one should examine it manually since automated methods will contain falls hits. However these can assist and reduce the time of investigation. It is advised to follow the best practice manuals for example from ENFSI[14] (the European Forensic Science Institute working groups) and The Scientific Working group for Image Technology and Digital Evidence to verify this.

There is good progress in deep learning algorithms, and at the Institute of Informatics they are widely used and optimized. Within the group Intelligent Sensory Information System led by Arnold Smeulders, where I cooperate with Marcel Worring we have a long history of collaboration. It started with a collaboration with Theo Gevers of pattern recognition of logos of drug pills. Later on we Arnold was in the scientific advisory board of the NFI, and there was a joint project MultiMedian where NFI was one of the partners, and many new developments in forensic video analysis were developed. In the meantime the group is award winning in the field of video analytics.

Deep learning

In the nineties we used the neural networks already for facial comparison and shoeprint comparisons. In the meantime the computing power also of GPU's expanded very much, so deep learning with more layers of a neural network was feasible.

Since the amount of multimedia data is growing rapidly, and UvA has already in collaboration with the Police and the Netherlands Forensic Institute developed software for handling this. One current project is the STW funded project Sort-it-Out were we also look on how to process the amount of data by using computer together with humans. The issue with humans is that we have a very good vision systems, however we are not very good in bulk processing of thousands of hours of video. The computer is better in processing large amounts of images.

Also caused by the calculation speed of current GPU board which makes it possible to learn an algorithms different concepts of images and the meaning of those. This works well for good quality images, however for most forensic images it is still a challenge[15] to the quality and since they are not standardized. In collaboration with Henri Bal of the Vrije Universiteit within the framework of the jungle computing[16] project that the calculation speed can be optimized by using the GPUs. Furthermore, optimizing the scheduling within Hansken is one of the research areas within this project.

If we look in casework, mostly our images are non-standard, and often the perpetrator tries to hide himself from being recognized or identified on images. Many perpetrators are aware of forensic traces and try to hide their identities behind masks and scarves when committing a crime, for instance a robbery. In practice one may also take care of look-alikes and family members that look more alike or even twins, which make the comparison more difficult.

Another issue is that we are very good in recognizing people that we know. However people that we do not know, that is something else. We see several experiments in literature on recognizing people that we do, know for example from Burton and others. Furthermore the question also arises if people that are trained are better in comparing than others. Many questions that are still not answered, and perhaps training a person or an algorithm within a case on hours of video will that help providing better answers? Also there is research in the United Kingdom at the Greenwich University[17] on the super recognizer[18], people that are much better in recognizing persons than the average person, and who are used often to look at hours of CCTV for intelligence purposes. Much more research also with larger standard datasets is needed to answer the questions.

Biometrics

We see much research in biometric processing, and here we also refer to the work of the research group in Twente of Didier Meuwly and Raymond Veldhuis. The collaboration of method for searching on concepts and combining this with text, location and other data seems to be important to reduce the manual search in big data. The graphical user interface and the way users interact with systems is of very important, and we see that the collaboration with the Technical University Eindhoven in the Sort-it-Out project tries to answer these questions.

If we look into forensic biometrics, we often have a case by case approach. The preselection is done at the police. For example if we have a comparison of a face, we would like to have reference material at the same position and same light as the suspect. If this is not possible, we will make reference material that is as close as possible to the light conditions and circumstances. I will go deeper into an example of casework where I am actually testifying for in court.

After that images are manually compared by 3 experienced independent examiners at the institute. They will compare features of the face and write their results on a standard form. The results of this is discussed and a consensus is reached. Also it is possible that no consensus is reached between the examiners and this is reported to the court.

The conclusion of the facial comparison is drawn based on a Bayesian framework first promoted to use in reports within the NFI by Marjan Sjerps and Charles Berger used at the Netherlands Forensic Institute, and is also recommended in a guideline by ENFSI. In this we might conclude for instance in a case that we have two hypothesis:

- Hypothesis 1: The suspect is the same person as the person on the CCTV-images of the robbery.
- Hypothesis 2: The suspect is another person with similar generic facial features as the person on the CCTV-images of the robbery (and is also not a direct family-member).

The Likelihood ratio is part of a standard verbal scale. This scale is used when does not have enough statistical information. The selection of the specific verbal term is based on expert knowledge, experience in research and casework. To enhance the understanding of the reader and the uniformity among the different experts, the NFI has defined the verbal terms numerically. These definitions are expressed in orders of magnitude and are listed in the right column in the table below. For example, the term 'slightly more probable' means that the probability of observing the results of the investigation is 2 to 10 times larger when one hypothesis is true than when the other hypothesis is true.

The conclusion expresses the evidential strength of the results regarding the hypotheses. The conclusion does not represent the probability that a particular hypothesis is true. That probability depends on other evidence and information outside the domain of forensic expertise and falls outside the scope of this report. More information about this way of concluding is available in the professional annex at the NFI website[19].

The findings of the investigation are much more probable if the suspect is the same person at the person on the CCTV-images than if it is another person with similar generic features The estimated Likelihood Ratio is much more than 1 and is estimated between 10000-one million.

The process mentioned to do this manually for many images is time consuming, however currently the methods to automate this are limited. Of course with good quality images this might differ, however as said often we have images that have less quality. Worldwide we see that in forensic investigations this manual method is still used by most labs. Automated procedures might help to determine how seldom a certain feature is that the examiner looks for. Also future generation of DNA-analysis can give us much more information on physical description of the face of a person and with this it is predicted that it could be possible to search or exclude persons.

Questions that arise are: can you search for the person in the orange coat that looks similar to our suspect. Can you track the person in the different videos and images? A we have seen in the past, the man-machine interface remains important here, since looking out thousand hours of video in combination with thousands of images that are collected from sources such as phones is time consuming. If one would search through millions or billions of images this is even more difficult.

Having said this, the innocence[20] project learned us that eyewitness misidentification was most of the cases, and after this invalidated forensics. For this reason methods used here should be used with consideration. Also earlier this year the FBI informed that in the nineties FBI Testimony on Microscopic Hair Analysis Contained Errors in at Least 90 Percent of Cases in Ongoing Review with DNA[21].

In the United states the National Academy of Science[22] reported that the Forensic science should be strengthened. This caused a major reform of forensic science, and many actions concerning quality assurance, bias prevention, determining source of error as well as education are implemented.

Since law systems in the Netherlands works differently, I expect the numbers are less in Netherlands. Also within the Netherlands Geert Jan Knoops, who will also give an inaugural lecture next week at this University, collected cases in the Netherlands for the innocence project[23]. Several cases were revised. One of the last cases with a major impact on the NFI was the Schiedam Parc Murder[24] in 2000 which led to many efforts in 2005 reduce the number of wrong convictions based on forensic evidence and also to improve the communication between NFI, court and police if doubts arise during the investigation. The Netherlands Register of Court Experts[25] led by Michel Smithuis is one of those efforts.

Hidden information: Camera identification

One research topic which had much interest is the camera identification[26, 27]. The camera identification or Pixel Response Non Uniformity is to find a link between a camera and an image. Due to imperfections during the fabrication of the image sensors, this is a kind of pattern in images that is normally not visible with the human eye. On the topic camera identification we have conducted much research with cameras and validation, and from the University of Applied Science of Amsterdam many students worked on this topic.

By using an algorithm[28] to filter this pattern out of a sequence of images or an image, the computer can fill a database of patterns. By using a correlation method of these patterns one can sometimes with a high likelihood ratio determine that the image has been made with a specific sensor. Validation in casework remains important, and procedures for using this in case reports are available. For intelligence software developed by the NFI, can be used to find a link, which should be investigated further.

This method can also be used for video, and for casework one can for instance find a link between a camera found at a suspect and child pornography images[29]. Also in other crime such as homicides this appeared to work as evidence to find a link between a suspect and a person that has been killed.

Even more one can also find relations between images if the camera is not in possession, as we can see with child pornography databases.

One issue is that comparing all these patterns is time consuming. So several approaches have been taken to limit the computation time. Also the use of GPUs is important in these, to find correlation between the data and here we worked together with the Vrije Universiteit to improve the method. Within the Netherlands Forensic Institute we developed software to do the forensic validation of camera identification. Also for intelligence we developed software to find relations between the different images, which can also cluster the images and find related images based.

In the past we developed several proficiency tests for different labs within the ENFSI framework to answer these questions for a closed sets of cameras based on the noise patterns. In these proficiency tests all labs have the right results on the comparison. In our group we report several cases a year on camera identification times in court as evidence.

The method also appears to work with social networks such as Youtube and Facebook, however recently it appears that the PRNU signal seems to be filtered out of the information. More research is needed to find what happened here. In practice it is possible to filter out the PRNU signal as we saw in the open source implementation PRNU decompare[30], which was developed by students of the group of Cees de Laat Software and Network Engineering of the University of Amsterdam.

Emerging biometrics Hands, feet and other biometrics

In child pornography cases, often the face is not visible. In this case we might do a forensic comparison of hands, feet and other body parts. These are conducted in a similar way as the facial comparison with three examiners comparing those independently from each other.

We are doing research on how well this works. Ana Slot, as student Master of Forensic Science worked on research in comparing hands with each other and this publication has been published in Journal of Forensic Science[31].

In the meantime we also have work on automated comparison implemented a framework for feet, which was further expanded for hands. We see that different features and deep learning approach will help in comparing these. The emerging biometrics are still in research phase, and can be used under standardized conditions.

Wearables and quantified self

If we look at wearables we see a large range of wearables, ranging from google glass, to watches and health devices. The wearables are often on a person for a long time, and streaming photos and video to cloud systems is a reality as much as the location information and what kind of actions someone did at a given time. We work on research on the digital evidence from google glass.

With the quantified self[32,33], we see that also heartbeat, breath patterns, blood pressure, walking, on a bicycle and other actions are collected which can also be helpful in a forensic investigation. The amount of data that people produce is growing, and of course a question remains if all this data can be used for forensic purposes based on privacy and security of the systems. They might even assist in investigations of the time of death.

Drones

The amount of drones[34] sold is increasing rapidly. These devices are important and challenges exist on what data can be taken from them. They can be also used for criminal acts, like attacking persons or dropping objects and intervene with regular air traffic. For that reason also data from these devices is important to collect in combination with time, date and extracting information from the owner of the device.

The drones are becoming much easier to use, since some of them have a follow me option so one does not have to control it anymore.

Sensors on mobile devices are becoming much cheaper, as we can for example see on the infrared sensors on mobile phones that are integrated easily on a phone for a few hundred euros instead of the thousands of euros they costed in the past. We also see that 3D imaging is getting cheaper nowadays, and the time is nearby that this is also integrated in smart phones[35].

Geo satellites

The development of higher resolution satellite imaging systems[36], which in future make maps from the Netherlands every few days, can be useful for investigation. The combination of all images including the geo satellites with the information of location can be very important in digital evidence, however investigators should be aware that uncertainty of location information and synchronization of time still remains an issue.

Disasters / large scale incidents

If we have large scale incidents and disasters, for example the MH17 investigation of an airplane, we see that there is lots of images and video collected from people that are in the neighborhood of the accident. All these images and video combined can be important to have a confirmation on what happened. This is one of the examples where Floris Gisolf of the Dutch Safety Board will do research on as a part time PhD-student.

With disasters also images might be used from the internet where the source is not exactly known. Manual manipulation detection is a time consuming work, and with separate images it is possible to conclude that the images have been manipulated. However if the ground source of the images is not available, determining that images have not been manipulated is more complicated to do. There are several tools and methods to automate this, however the issue is also here, once there is a detector published, one can circumvent the method by taking care of the detection method. Also we see that the automated detectors in standard sets of manipulated images only score with 90 percent at best. Depending on the use, this can be acceptable, however this is a factor one should be aware of.

Combining patterns

Within the field of forensic data science one should combine patterns. As we saw with the research of Peter Sloot and Paul Duijn[37] with simulation approaches are becoming feasible also to work on predictive methods based on big data and mining. However, as these criminal networks operate in secrecy, data-driven knowledge concerning the effectiveness of different criminal network disruption strategies is very limited. By combining computational modeling and social network analysis with unique criminal network intelligence data from the Dutch Police, they discovered, in contrast to common belief, that criminal networks might even become 'stronger', after targeted attacks. For the PhD-thesis of Paul Duijn I am one of the promotors. As editor of a special issue of the journal of Digital Investigation as special issue will be published within a month, on the big data and intelligent data analysis.

Within the Institute of Informatics, several collaborations are foreseen with the Information and Language Processing group of Maarten de Rijke. Currently a NWO project on Semantic E-search is funded, and NFI is one of the partners in the project, which provides methods that are very useful in forensic practice. In future other proposals within this field are in preparation.

Also with the group Software and Network Engineering (SNE) of Cees de Laat we have much collaboration. Within the Master SNE, the track digital forensics the NFI provides lectures. Several student forensic research projects with some of them results that could be directly used as research. Also it is foreseen that we might submit joint research proposals in future.

The question with the implementation of the combined approach of patterns is does the court still understand the visualization and reporting that we write. We sometimes have the chance to present our work in the court, and can clarify the results. There are several approaches within the Netherlands Forensic Institute also with judges to learn from each other how to present this in a more understandable way, and perhaps also help them to determine priors in a case. We see that the knowledge in courts on the methods is expanding due to training and courses at the court, however the question is still remains we really understand each other or do we think that we understand each other?

When I refer to my first appearance in court, it was with the Ballpoint case in 1995, now twenty years ago. This was a case where a woman was killed with a ballpoint through her eye. One of the hypothesis was she was murdered with a crossbow and a ballpoint. In that time I was active within ballistics and when the prosecutor was summarizing the findings of my report at the end, and it really was not the way I presented it or had made the conclusions in the report, so I decided to inform the court on this during the trial, which was an unusual experience. Since that many improvements have been made also in reporting and training of both the courts and the experts at the NFI. However it remains important to check that the end users of the reports of the NFI understand what is written, and also the experts at the NFI understand the questions from the court and the police. In twenty years I hope we can look back to many developments that have assisted us to improve the understanding of the use of big multimedia data.

The issue of language use and combining different statistical approaches should be also within complex crime scenes one of the areas where we need continuous improvement. Validating the methods and the presentation in court remains important, and also with the exponential rate of change of digital techniques that are implemented in society.

Dankwoord

Graag sluit ik af met een woord van dank. Allereerst dank aan het College van Bestuur van de Universiteit, de decaan van de Faculteit Natuurwetenschappen, Wiskunde en Informatica Karin Maex, en aan de leden van het bestuur van de Stichting Leerstoel Criminalistiek en het Curatorium, voor het in mij gestelde vertrouwen. De directie van het Nederlands Forensisch Instituut in het bijzonder Reinout Woichiez, Kees Mohring en Tjark Tjin-a-Tsoi, die zich via de Stichting Leerstoel Criminalistiek persoonlijk hebben ingezet deze leerstoel forensic data science. Ik dank Jan Bergstra, Arnold Smeulders en Marcel Worring voor hun inzet om deze leerstoel bij het Instituut voor Informatica te realiseren. Bij het curatorium wil ik danken in het bijzonder Jose van Dijk, Maarten de Rijke. Ook dank ik Arian van Asten voor en Maurice Aalders, directeuren van het Co van Ledden Hulsebosch Centrum voor hun adviezen en inzet.

Veel dank aan mijn promotor Max Viergever en co-promotor Jurrien Bijhold, die beiden hier aanwezig zijn. Verder wil ik danken de afdelingshoofden Petula Huising en Erica Rietveld, beiden om hun inspirerend leiderschap.

Ik wil danken mijn naaste collega's Arnout Ruifrok, Rikkert Zoun, Arjan Mieremet, Derk Vrijdag, Daniel Paloma van Es, Henk Leijenhorst, Marjolijn Brouwer, Linda Dekkers, Lotte Smelik, Menno Israel, Didier Meuwly, Marjan Sjerps, Ingrid Bosman, Ivo Alberink, Jeroen van den Bos and Cor Veenman for their support, and all colleagues from the departement Digital Technology and Biometrics.

Bij de UvA wil ik danken Peter Sloot, Andrea Haker, Hamideh Afsarmanesh, Alfons Hoekstra, Cees de Laat, Stefan Rudinac, Jan Zahalka, Bert van Es, Max Welling en uiteraard alle andere collega's die ik vergeet te noemen. Bij de VU Amsterdam wil ik danken Henri Bal, en de collega's van het e-Science Center in het project. Als oud collega's wil ik graag danken Wim Neuteboom, Bert van Leuven, Sijtze Wiersma, Jan Keijzer, Hans Henseler, Huub Hardy, Menno van der Marel en Ronald Prins.

Van de politie wil ik danken Wim Michels, voorzitter van het Hoofden Digitale Expertise Overleg, evenals Barend Frans, Frans Kolkman, Ruud Elderhorst, Eric Loermans, Chris Oornick, Peter Duin en Ron Boelsma. Bij de FIOD wil ik danken Terry Hartsinck. Bij de Koninklijke Marechaussee Gerard Looijen. Bij het Ministerie van Justitie en Veiligheid dank ik Arno van Oosterhout, Bob Stap en Nicky Waterreus voor de goede gesprekken en vele adviezen. Verder dank ik alle collega's die ik niet bij naam mag noemen.

Ik wil danken vanuit ENFSI, Jan de Kinder voorzitter van de ENFSI die deze leerstoel heeft gesteund, evenals Mikael Lindstrom van Europol, en vice-voorzitter van de werkgroep Forensische ICT waar ik voorzitter van ben.

Ook wil ik mijn familie bedanken zonder wie ik dit nooit zou hebben bereikt. Allereerst mijn echtgenoot Lammert Brons, die me in voor- en tegenspoed altijd steunt met het werken hierin, en ik ben blij te zien dat zijn ouders hier ook zijn. Hij is werkzaam als teamleider van het Leger des Heils, en via deze wijze zien we de verschillende kanten van de maatschappij. Iemand kan vanuit de reclassering verder weer de maatschappij in worden geholpen, maar ook cliënten die een strafbaar feit toegeven zonder dat ze het gedaan hebben, waarbij hopelijk forensisch bewijs aan toont dat er een andere dader moet zijn. Ik ben enorm gelukkig met mijn lieve Lammert en eigenlijk zou het dankwoord voor jou natuurlijk het langst moeten zijn.

Verder wil ik mijn moeder die hier is, en dank haar voor haar geduld en toewijding, broer en zus bedanken voor al hun steun, evenals mijn wijlen vader. Gelukkig is de zus van mijn vader, Alice Geradts, hier aanwezig en ook haar dank ik voor haar enthousiasme. Als herinnering aan mijn vader heb ik de toga van binnen laten bekleden met stof met de afbeelding van Havank. Havank was zijn favoriete detectiveroman, en is ook de naam van het vingersporen-identificatiesysteem in Nederland. Ik dank u voor uw komst en aandacht.

Ik heb gezegd.

Noten

1. Kurzweil, Ray. *The law of accelerating returns*. Springer Berlin Heidelberg, 2004.
2. “What Is Big Data? – Gartner IT Glossary – Big Data.” 2013. 28 Oct. 2015 <http://www.gartner.com/it-glossary/big-data>
3. “Hype Cycle Research Methodology | Gartner Inc.” 2010. 23 Nov. 2015 <http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp>
4. “Eight (No, Nine!) Problems With Big Data – The New York ...” 2014. 23 Nov. 2015 <http://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html>
5. Lazer, D. “The Parable of Google Flu: Traps in Big Data Analysis.” 2014. <https://www.sciencemag.org/content/343/6176/1203>
6. Davenport, Thomas H, Paul Barth, and Randy Bean. “How ‘big data’is different.” *MIT Sloan Management Review* 54.1 (2013).
7. van Baar, RB, HMA van Beek, and EJ van Eijk. “Digital Forensics as a Service: A game changer.” *Digital Investigation* 11 (2014): S54-S62.
8. van Beek, HMA et al. “Digital Forensics as a Service: Game On.”
9. “13-132 Riley v. California (06/25/2014) – Supreme Court.” 2014. 30 Oct. 2015 <http://www.supremecourt.gov/opinions/13pdf/13-132_8l9c.pdf>
10. “Digital forensics resembles the Wild West when it comes to ...” 2015. 23 Nov. 2015 <http://www.techrepublic.com/article/digital-forensics-resembles-the-wild-west-when-it-comes-to-regulation/>
11. “Casey Anthony Trial Fast Facts – CNN.com.” 2013. 30 Oct. 2015 <http://www.cnn.com/2013/11/04/us/casey-anthony-trial-fast-facts/>
12. “Best Practice Manual of Forensic Examination of Digital Technology” http://www.enfsi.eu/sites/default/files/documents/enfsi-bpm-fit-01_2.pdf
13. “ENFSI | Home.” 2006. 28 Oct. 2015 <http://www.enfsi.eu/>
14. “ENFSI | Home.” 2006. 24 Nov. 2015 <http://www.enfsi.eu/>
15. “Sort It Out: Visual Analytics for Multimedia Collections | STW.” 2014. 28 Oct. 2015 <http://www.stw.nl/en/node/6973>
16. Seinstra, Frank J et al. “Jungle computing: Distributed supercomputing beyond clusters, grids, and clouds.” *Grids, Clouds and Virtualization* (2011): 167-197.
17. “Take this test to see if you’re a super recogniser – part of an ...” 2015. 28 Oct. 2015 <http://www.gloucestershireecho.co.uk/test-super-recogniser-elite-group-forget-faces/story-26705528-detail/story.html>
18. Russell, Richard, Brad Duchaine, and Ken Nakayama. “Super-recognizers: People with extraordinary face recognition ability.” *Psychonomic bulletin & review* 16.2 (2009): 252-257.
19. http://www.forensischinstituut.nl
20. “Home — The Innocence Project.” 2002. 24 Nov. 2015 <http://www.innocence-project.org/>
21. “FBI admits flaws in hair analysis over decades – The ...” 2015. 24 Nov. 2015 <https://www.washingtonpost.com/local/crime/fbi-overstated-forensic-hair-matches-in-nearly-all-criminal-trials-for-decades/2015/04/18/39c8d8c6-e515-11e4-b510-962fcfabc310_story.html>

22. Roux, Claude, Frank Crispino, and Olivier Ribaux. "From forensics to forensic science." *Current Issues Crim. Just.* 24 (2012): 7.
23. Cases – Knoops' advocaten. 2012. 4 Dec. 2015 <http://www.knoops.info/en/knoops-innocence-project_en/cases>
24. van Koppen, PJ. "Blundering Justice – Springer." 2008. <http://link.springer.com/chapter/10.1007%2F978-1-60327-049-6_12>
25. "Nederlands Register Gerechtelijk Deskundigen (NRGD ..." 2008. 8 Nov. 2015 <http://www.nrgd.nl/>
26. van Houten, Wiger, and Zeno Geradts. "Using anisotropic diffusion for efficient extraction of sensor noise in camera identification." *Journal of forensic sciences* 57.2 (2012): 521-527.
27. Gisolf, Floris et al. "Common source identification of images in large databases." *Forensic science international* 244 (2014): 222-230.
28. Gisolf, Floris et al. "Improving source camera identification using a simplified total variation based noise removal algorithm." *Digital Investigation* 10.3 (2013): 207-214.
29. Gisolf, Floris et al. "Common source identification of images in large databases." *Forensic science international* 244 (2014): 222-230.
30. "PRNU Decompare download | SourceForge.net." 2009. 28 Oct. 2015 <http://sourceforge.net/projects/prnudecompare/>
31. Slot, Ana, and Zeno JMH Geradts. "The Possibilities and Limitations of Forensic Hand Comparison." *Journal of forensic sciences* 59.6 (2014): 1559-1567.
32. Swan, Melanie. "Sensor mania! the internet of things, wearable computing, objective metrics, and the quantified self 2.0." *Journal of Sensor and Actuator Networks* 1.3 (2012): 217-253.
33. Swan, Melanie. "The quantified self: Fundamental disruption in big data science and biological discovery." *Big Data* 1.2 (2013): 85-99.
34. Stuckenberg, David, and Stephen Maddox. "Drones in the US National Airspace System." *International Journal of Aviation Systems, Operations and Training (IJASOT)* 1.2 (2014): 1-22.
35. Akaoka, Eric, Tim Ginn, and Roel Vertegaal. "DisplayObjects: prototyping functional physical interfaces on 3d styrofoam, paper or cardboard models." *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction* 24 Jan. 2010: 49-56.
36. Kushwaha, Deendayal, Sridhar Janagam, and Neeta Trivedi. "Compute-Efficient Geo-Localization of Targets from UAV Videos: Real-Time Processing in Unknown Territory." *International Journal of Applied Geospatial Research (IJAGR)* 5.3 (2014): 36-48.
37. Duijn, Paul AC, Victor Kashirin, and Peter MA Sloot. "The relative ineffectiveness of criminal network disruption." *Scientific reports* 4 (2014).