Document Fraud Detection at the Border

Preliminary observations on human and machine performance

Monica Gariup
Frontex, Research and Development Unit
Warsaw, Poland
Monica.gariup@frontex.europa.eu

Gustav Soederlind
Frontex Research and Development Unit until March 2013
Göteborg, Sweden
Gustav.soderlind@gmail.com

Abstract—How many false documents (forged and counterfeited) cross the border undetected? What is the real extent of the phenomenon of document fraud at the border? How good are border officers and their technical equipment at detecting fraud in the first line? How can we measure the impact of capacity-building measures (training and technology) in the field of document fraud detection? This paper proposes to approach the traditional problem of “known unknown” of risk analysis by taking the performance of detection capabilities – human and machine-supported – seriously. It argues that capability-based vulnerabilities need to be systematically assessed quantitatively and qualitatively in order to make sense of the risk and to devise, test, and measure the effectiveness of countermeasures. The paper reports the preliminary results of an exercise simulating the first line of document inspection at the border. European document experts and automated document inspection systems were challenged to recognize genuine and false documents under a very tight time constraint. Although the experiment suffered of many methodological weaknesses due to the limitations of the context in which it was conducted, a number of initial observations can be drawn on the importance of human skills and experience; the strengths and shortcomings of automated systems; and the need to further test and study how human and machine capabilities can be improved and combined in order to increase their detection effectiveness and thus strengthen border security.

Keywords Automated Document Inspection System, Document Fraud, Electronic Passports, Border Control, Human-Machine Interaction, Risk Analysis

I. INTRODUCTION

The verification of the authenticity of a travel document is essential in order to establish the identity of a person crossing a border. It is estimated that in the European Union some 23,000 instances of document fraud (forgeries and counterfeiting) are detected at border control per year [1]. Considering the overall border traffic at the Schengen borders however (some 800 million border crossings per year), this means that either the problem of document fraud is statistically insignificant – and thus may affect risk prioritization, or that the real problem is the lack of detection or a possibly serious internal capability-based vulnerability that needs to be remediated.

Since detection numbers/data constitute the only quantitative source to describe the phenomenon (in terms of volume, nationality of documents and users, routes, detection points etc.) or put in other words define what the problem is, this paper argues that attention needs to be paid on how these false documents are detected (detection point triggered by document inspection by the officer, by the equipment, passenger profiling etc.) in order to better assess whether the detection numbers provide a realistic situational picture at the border, to spot potential shortcomings and vulnerabilities, and to identify what capacity-building measures, if any, need to be introduced and tested to strengthen border security.

First line inspection of the document and of its holder determines whether suspicion exists and further controls in the second line are necessary. In an attempt to counter the phenomenon of document fraud, however, the physical, optical and electronic security of travel documents (passports and visas) has become so complex that a border control officer often needs to rely on the support of automated document inspection systems (including register checks) to perform this task in a timely and accurate manner. This reliance on the machine is often cause of concern for experienced officers who trust their senses and experience more and who fear automation will make them and their professional profile redundant.

To better understand what triggers detection and how effective detection is, it is thus important to start isolating the variables – the officers, the equipment, their interaction - and to observe/assess them in an operational environment. Unfortunately, because of the traditional sensitivity related to the effectiveness of law enforcement and/or lack of resources, data on how well detection is performed – especially when this involves the evaluation of the performance of officers or of equipment for which a considerable investment was made – is scarce or not publicly available. In addition, as it is the case for all research that involves human behavioral components and individual subjectivity, it is often difficult to collect and reliably generalize data on how decisions were made in the first line and to disaggregate what triggers the officer’s suspicion in a real operational environment.

In order to stimulate a discussion on the detection challenge in the first line and to learn from practice how to refine the research methodology, Frontex organized in June 2012 together with the National Document Fraud Unit of the United Kingdom and with the support of the Border Security Training Centre of the Dutch Royal Marechaussee a one-day exercise that saw 26 European police document experts and 10
document inspection systems confronting the task to determine whether 104 real passports were genuine or false in an average of 15 seconds per document. This paper summarizes the results, observations, and lessons learned during the exercise. Despite the theoretical ambition of the research rationale explained above, the exercise should be seen as an initial trial attempt to better understand how to approach the research question rather than a scientifically and methodologically sound experiment.

After a short overview on the background and the state of the art of existing research on document fraud detection capabilities, the paper presents the central theoretical concepts and the methodology used. Major results and observations are discussed as well as lessons learned for further testing and research.

II. RESEARCH BACKGROUND AND RATIONALE

The Frontex Annual Risk Analysis (ARA 2012) released in April 2012 identified document fraud as a high priority issue [1]. In particular, the report noted the following challenges:

- Increasing sophistication of document falsification/counterfeiting;
- New techniques to circumvent biometric checks;
- Use of genuine documents by imposters (impersonation, stolen identities);
- The abuse of genuine supporting documents to obtain genuine travel documents (abuse of legal channels and fraudulently obtained documents);
- The abuse of supporting documents to justify purpose of stay;
- Number of detected document fraudsters increasingly under-represents the overall extent of the phenomenon;
- Fraud expected mostly among EU travel documents.

In addition, ARA2012 identified the risk of less effective border control due to more reliance on technological equipment (triggered by reduced resources and increased passenger flows) and a decrease in the importance of human interventions that may increase the risk that a new modus operandi passes undetected and reduces the possibility to collect human intelligence.

This analytical picture is supported by the observation of operational practices at several Schengen Border Crossing Points (BCPs) gained during Frontex organized Joint Operations, questions raised while studying Automated Border Control (ABC) Systems using electronic passports as deployed at several airports in the Schengen area, and discussions with border control officers about the relevance of detection numbers, their relationship with technology, and their perception of effectiveness, problems, and challenges in performing their tasks.

Such a state of affairs prompted the research need to advance the understanding on what variables affect the decision of first line officers, and how accurate the entry decision is taking into consideration both technology and the human factor so as to contribute to the development and improvement of human and technical decision-support capabilities (and justify further investments).

III. STATE OF THE ART ON DOCUMENT FRAUD DETECTION CAPABILITIES

Existing publicly available research in the field of the effectiveness and performance of document fraud detection is scarce. When it exists and is published, it tends to focus on the technical equipment. Because of the sensitivity of performance measures in the world of law enforcement, collecting and analyzing data in the field and being able to publish them is a particularly daunting endeavor. None the less, the few attempts at studying the issue at stake have proven a valuable background to this research.

A. Interoperability Testing of E-Passports and Implicit Consideration of Inspection Systems

Since ICAO with Doc. 9303 has established the standards for Machine Readable Travel Documents (MRTDs), a number of interoperability and usability testing rounds have taken place all over the world. These exercises have been mainly designed to test whether the issued travel documents really follow the technical standards set out in the ICAO document and, as a consequence, whether they are interoperable and can thus be read by readers in different countries of the world. Such testing has been instrumental to highlight not only the lack of compliance to the standards because of technical difficulties and faulted interpretation of the specs – now greatly diminished since the early 2000s – but, of importance here, the great variation in reader performance [2].


B. Research conducted by European governmental authorities

Some European national border control authorities (the Netherlands, Portugal, Germany for instance) have conducted studies into the effectiveness of the equipment used – often in the context of technical testing related to procurement and

2 For example, “the Department of Homeland Security's first tests of electronic-passport interoperability exposed technology flaws, including myopic and dyslexic smart-card readers. Some readers could not detect the presence of e-passport chips, many could detect the chips but could not read them and others were befuddled about what information they were supposed to display” [2]. In Europe, Council Regulation (EC) No 2252/2004 established standards for security features and biometrics in passports and travel documents issued by Member States. The now defunct Brussels Interoperability Group (BIG) was the technical group working under the authority of the Article-6-Committee of the European Commission, responsible for technical interoperability of MRTD and Residence Permits in Europe. Its prime mission was to act as the focal point for resolving all technical issues that arose from the development, implementation, testing and application of fingerprints/EAC in MRTDs including a Certificate Policy. During its life BIG contributed to organize interoperability testing in different locations in Europe.

1 The Frontex Risk Analysis products are based on data and analytical input regularly provided by EU Member States.
acquisition. Although the studies themselves are confidential, their major conclusions can be summarized here.

The Dutch ECID Advies & Innovatie (Expert Center for Identity Fraud and Documents) tested in 2008 and 2009 several document readers. The research – unpublished and restricted - focused not solely on document verification, that is, properly recognizing (the correct model) of passports with a machine readable strip and the eventually present RFID-chip, but in particular on correctly recognizing forgeries of these passports based on the security features found within these passports. The aim was to put emphasis on the need to maximize the occurrence of true negatives and minimize the occurrence of false positives. Previous tests in 2008 had in fact revealed that all participating suppliers chose to maximize true positives while minimizing false negatives with debatable thresholds, thus giving preference to speed at the cost of security. The tests proved that, despite some weaknesses, E-passport readers are suitable for reading and controlling the machine readable zone (MRZ) and the RFID-chip. However, it was also noted that important information contained on pages different than the first biographical page – such as visas and changes thereof, stamps, extension of expiration dates etc. – could not be captured and its authenticity verified.

In November 2010, Portugal presented a Note to the Working Party on Frontiers/False Documents (Council of the European Union), in which it discussed results of a preliminary study on the reasons why detection of false documents seemed to be declining and on issues related to the performance of automated document examination equipment. Although these observations are not supported by empirical statistical evidence, they support the need to better understand what affects the detection and the role of capabilities (humans and equipment) in that relationship. The Note acknowledges the strengths of the advanced functionalities of document readers and the reinforced security of contemporary travel documents. However, it also reports an overreliance on the automated result of the readers, the fact that template databases used are often inaccurate, and the inherent opacity for the human of electronic security mechanisms. All of these “often mislead[s] immigration officers to the point of neglecting their own instinct”; and “[p]otentially significant numbers of documents are accepted as genuine when in fact they should be considered suspect or questioned”, thus concluding that “the request for (more) facilitation compromises better performance”. The report concludes that “according to the Portuguese experience, it is not possible to determine whether the fall in fraud detection rates is related to the use of the readers, in the same way as we may not determine if such reduction is related to the improvement of the physical and logical security features of travel documents. The implementation of automated components at the borders may result in a less detailed checking procedure or, in other words, to an over confident attitude towards the checking made by the equipment”. [4] 

The German Bundeskriminalamt has been testing how automated document readers perform when checking optical security features. Performance of devices varies, and the main conclusion – in a study still unpublished and restricted - suggests that the design of secure documents (passports) needs to take into consideration the fact that inspection systems may not be able to verify the added security feature, thus making the newly issued document not necessarily more secure (because the feature is not recognized by the reader or the officer).

Frontex study on E-passports highlights among other things the perceived technical risks inherent to inspection systems (misconfiguration, attacks, failure etc.). More importantly for this paper, it stresses the need for enhanced training of border officers in the use and understanding of the limitation of the technology since misinterpretation of errors may constitute a vulnerability in itself [6].

IV. KEY CONCEPTS

A. Definition of the Detection Task

The Schengen Handbook states that document examination in the first line “consists of a rapid and straightforward verification of the travel document, with the aim of checking the validity of it and of detecting the presence of signs of falsification or counterfeiting (point 14, page 10)” (such as photograph/image substitution, page substitution, alteration of personal details, alteration or counterfeiting of stamps or visas or the complete reproduction of a document etc.) [7]. The examination, as described in the Schengen Catalogue, should be supported with equipment like:

- terminal for consulting SIS, VIS and national databases
- portable terminal when necessary;
- document examination instrument, with variable UV light, white overhead light, transmitted light;
- magnifying glass at least 10 x or mono-microscope with variable zoom;
- retro-reflective lamp;
- Schengen entry and exit stamps;
- Schengen Borders Code and Schengen Handbook with annexes;
- constantly updated risk indicators and risk profiles;
- available electronic document examples and other information necessary for carrying out border checks [8].

In addition, “integrated passport readers should be used at first-line checks (mobile and fixed readers) at Border Crossing Points with extensive traffic” as well as “automatic consultation of SIS, VIS and relevant national databases for third country nationals” [8].

For what concerns human capabilities, border control officers in the first line should possess the following skills:

- Ability to distinguish colors, shapes, depth, and to observe details;
- Awareness of different characteristics and components of security documents: ICAO standards, substrates, inks, assembly, production methods and bio-data
- Awareness of overall threats and relevant intelligence information;

3 See also [5]. In particular, the level of skills and specialized equipment of border guards and police officers has an important influence on the capacity of law enforcement personnel to detect counterfeit documents in the first place.
Human Factor research are for instance:

- Knowledge of the various types of primary security features (e.g. watermarks, optical variable inks);
- Knowledge of the different kinds of fraudulent documents (how they are falsified and counterfeited);
- Knowledge and ability to use available equipment commensurate with training;
- Ability to detect and identify the various types of primary security features [9].

No specific guidelines exist in order to test and assess the performance of these capabilities in an operational setting.

**B. Human Factors Research, Human-Machine Interaction, and User Cases for Document Examination**

Because of the increase in the use of automation to perform border control functions and the potential risks that this may introduce from a security perspective, concepts and research in the area of Human Factors, Decision-Making and Human-Machine Interaction offer interesting insights.

The application of Paul Fitts’ Model on Human and Machine Excellence, developed initially to understand how to allocate tasks when using computers in aviation security, is useful to discuss the strengths and weaknesses of each document examination use-case but also in reference to the overall decision-making process in the first line [10]. In particular, it helps critically question and address a number of issues related to function allocation and the role of decision-aids that the relative new intensive use of technology in the first line raise.

Some still unanswered questions raised by referring to Human Factor research are for instance:

- What tasks should border officers perform, and what tasks machines should perform?
- How can human and machine cooperate in an efficient way? (see issue of trust and confidence in the machine (over-reliance/ mistrust)?
- How to distribute functionalities if the error rate – performance limitations (of the human) – Human Error Probability (HEP) and False Acceptance Rate (FAR) – and of the Machine are not known?
- What affects the performance of the humans (physiological - psychological)?
- How to ensure that the human is in full control of the decision he/she needs to take (issue of human motivation)?
- How should the border control officer (the HUMAN) be trained for the new tasks including the use of complex equipment (attention, cognition, workload, decision-making)?
- How to mitigate the new (and old) risks introduced in the process (Attempts to defeat the machine/the human)?

<table>
<thead>
<tr>
<th>USE CASES</th>
<th>STRENGTHS</th>
<th>WEAKNESSES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. HUMAN ONLY</strong></td>
<td>Manual by border control officers without the use of technical support</td>
<td>Physical (including booklet and stamps)</td>
</tr>
<tr>
<td><strong>2. HUMAN - MACHINE</strong></td>
<td>Manual by border control officers with the support of computers, optical readers, and watch-lists</td>
<td>Physical Optical Databases</td>
</tr>
<tr>
<td><strong>3. HUMAN - MACHINE ++</strong></td>
<td>Manual by border control officers with the support of computers, optical readers, checks against watch-lists, and infrastructure for reading electronic passports</td>
<td>Physical Optical Electronic Databases Biometrics possible</td>
</tr>
<tr>
<td><strong>4. MACHINE ONLY</strong></td>
<td>Automated gates for document authentication and biometric verification</td>
<td>Optical Electronic Databases Biometrics Speed</td>
</tr>
</tbody>
</table>

*Table 1 Use Cases for Document Examination in the First Line of Border Checks*

Each use-case applies differently depending on the BCP context and setting. However, due to the increase in the sophistication of travel documents from a physical, optical, and electronic security perspective, the trend has been on the predominance of the use-case here called “human-machine+++” (or “hybrid human-machine system”) that tries to combine the advantages of the human “smell, touch, and feel” skills with the superior speed, sensorial sensitivity and computational power of the machines (technical equipment). The reality is simply that the HUMAN ONLY use-case, although still used especially as a fall-back procedure in case of failure of the machines, contains intrinsic vulnerabilities related to the lack of optical and electronic sensorial abilities of the humans (for instance, the officer would not be able to open a chip, confirm the authenticity, and read its content by using its human abilities alone and training in this regard is not of much help). The same applies to a lesser extent for the case of MACHINE ONLY where the important factors of human judgment, flexibility, and improvisation and so on are eliminated from the process.

**V. THE DOCUMENT CHALLENGE EXERCISE**

In trying to address some of these questions, the Document Challenge Exercise’s initial objective was to try to focus on the HUMAN ONLY use-case and the MACHINE ONLY use-case (without databases and biometrics) in order to better understand the strengths and weaknesses of both use-cases and identify the areas in which combination is not only necessary but essential. As it will be seen in the discussion of the results of the exercise, this attempt failed because of methodological flaws inherent to the type of the experiment performed.

The exercise aimed at simulating document examination in the first line of border control in order to better understand:
a) How (method) and how well (performance) human experts verify document authenticity (physical and optical) under time constraint and without significant technical help;

b) How (method) and how well (performance) automated document inspection systems (also here referred to as “machines”) verify document authenticity (physical and optical) under time constraints;

c) Whether machines significantly outperform human experts;

d) What type of errors do humans and machines make;

e) What weaknesses/vulnerabilities are intrinsic to human and machine document verification taken in isolation and how these vulnerabilities can be mitigated in the future?

It should be noted that initially the “challenge” was meant only to bring in industry to present its products and to raise risk awareness as an addendum to a meeting organized to discuss training issue. It quickly developed into a more demanding exercise that, because of lack of time and lack of control on both participants and data-sets, was not based on a fully developed research protocol. In line with the approach “learning by doing” of Action Research, the exercise was designed by and for practitioners in order to cooperatively identify and address issues that affect the overall “community of practice” (officers but also technology providers) involved in border control [11].

A. Methodology and Participants

The exercise was performed on 28 June 2012 at the Border Security Training Centre in Schiphol, Amsterdam (the Netherlands) in collaboration with the UK National Document Fraud Unit (NDFU) and the Frontex Training Unit (TRU).

26 Members of the Frontex Document Specialist Board (DSB)\(^4\) participated in the exercise as well as 10 Document Reader providers/integrators who answered an open invitation to participate. The document experts and the equipment (operated by the vendors) were challenged to correctly detect genuine and false documents with an average time of 15 seconds per document using a test-set of 104 genuine and false passports. The exercise focused on document examination only: other essential tasks performed during a first line border check (such as profiling, registry checks and identity verification) were deliberately excluded. It should be noted that this is a very artificial de-coupling since in operational practice the document is never considered in isolation from other variables that affect the overall entry decision. This also means that the time allocated per passport examination (15 seconds per document) is also very artificial: in reality in this very short time first-line officers are called to perform different activities in parallel with relative high demands in cognitive efforts, attention, and accuracy.

In order to allow for a comparison of the human and the machine performance and because of lack of time and resources available, the following limitations were introduced:

- Same amount of time pressure of approximately 15 seconds per document (includes the actual examination of the document and the time needed to quickly report on the result in writing) for both human and machine examination;
- Central focus on optical security features that are readily available for control also by the human eye;
- Genuine/false documents decisions were collected by the participants using a uniform form.

Although most of the passports were provided by the UK and the Netherlands, it was made sure that the type and nationality of the passports, and the types of forgeries were as distributed as possible in order to give all the experts a sense of the possible diversity in documents in circulation. This is because experts especially from smaller EU Member States and BCPs may encounter only few examples of certain types of forgeries or passport type and become specialized in a certain modus operandi rather than be able to recognize forgeries independently of their individual operational experience. In total the exercise contained 19 genuine out of 104 documents in total.

A slight majority of passports (62 out of 104) were from EU Member States. A correlation between expert nationality, passport flows at specific BCPs, and accuracy rate could be done in order to better understand the strengths (and weaknesses) that a localized operational experience can bring to the formation of a document expert. However, such an analysis was not performed in this research due to the limited number of documents and experts participating.

The document specialists are national experts in the field of document fraud formally assigned by their respective national border management authorities to assist the Frontex Training Unit in shaping the content and format of new trainings offered by Frontex in this domain. An important prerequisite is expert knowledge in the domain of secure documents, albeit not necessarily acquired in the law-enforcement domain. Experts participating in the exercise were representatives of the following countries: Austria, Belgium, Bulgaria, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom. It is important to note that the background and experience of the experts is directly linked to their performance and the accuracy in detection. The exercise should have been conducted – and maybe should be replicated in the future – with real first line border officers. For logistical reasons, however, this was not possible and represents an important limitation.

Most of the experts were handling travel documents during their normal duties (in second or third line) and had extensive document expertise. Only 46% of the participants had first line experience with an average of 7.5 years.

The following providers answered to a public call for participation and took part in the exercise: Inspect & Detect BV; AVD Solutions.Sgraba; BPI; AU10TIX LTD; Foster & Freeman Ltd.; Securitech Detection & Verification; Trusted Terminal Ltd; ID Scan Biometrics; Bundesdruckerei GmbH;

\(^4\) The Frontex Document Specialist Board is composed of document experts designated by the Member States. The Board meets regularly to advise the Frontex Training Unit on curricular activities.
and Keesing Reference Systems. Two of the devices did not match the profile Frontex had expected, since the aim was to observe how equipment performed automated decisions based on comparisons of the scanned data-page against the matching images/templates in a reference database. The two outliers were nevertheless accepted as they could provide a reference for comparisons with the other equipment and would add value to the equipment demonstration for the document experts. Technical specifications of the devices that participated in the exercise were collected.

Surprisingly and a factor the research team could not control, 30% of the machine operators had law enforcement and first line border control experience (average of 4 years). This obviously introduced a spurious element in the exercise and the need to revise the results and recognize that the human factor may have played a role in the performance of the machine as well.

Each participant was asked to report their decision in the format genuine/false. In addition, they were asked to note detection point and confidence level. However, because the time available for the exercise was very limited, only the information on genuine/false was considered reliable and usable for analysis.

B. Major Results and observations

Results, although limited and statistically non-representative provided interesting insights. The detection rate of the human experts was used as a baseline from which to evaluate the accuracy of the document readers

1) Human Experts

The performance of document experts in correctly identifying genuine and false documents was in general surprisingly accurate especially considering the time pressure they were not used to (since they mostly deal with documents in the second and third line).

Interestingly, it seems that genuine documents were the most difficult to identify since many experts wrongly categorized them as false. In fact, when ranking the documents in order of difficulty (i.e. the least number of correct answers per document) the top 15 documents were all genuine (note that there were only 19 genuine documents in the test-set). An assumption could be made that the document experts have a bias against genuine documents as the documents that second line officers encounter during their work are in many cases already flagged as suspicious, and thus the proportion of genuine versus fraudulent documents that they see varies significantly from what first line officers experience. It raises the question of whether first line officers could also have a bias, albeit in the opposite direction. First line experience did not however significantly affect the results.

2) Document Inspection Systems

The document readers tended to do less well than the document specialists and also managed to check fewer of the documents in the given time.

The experts generally easily detected documents in which:
- Watermarks were missing or poorly simulated;
- Inkjet had been used for the background, sometimes quite poorly;
- Tactile features were missing;
- Evidence of manipulation was “obvious” by looking at it, e.g. the outline of the new/old image was visible.

In general however, apart from device number 3 that was not fully automated (what explains the relative high number of passports not examined in the given time) and device number 10 that suffered from numerous software reboots, no major outliers could be identified in terms of performance.
It should be noted that there are 3 independent equipment specific aspects that may impact on the performance of the document reader:

- The capability of the scanner, e.g. resolution, quality and position of lighting, speed of scan;
- The image template database, e.g. number of missing documents, selection of security features in available documents;
- The Software: decision logic, speed, presentation, thresholds.

In addition there is operator training and expertise, which becomes extra important when the results require evaluation of multiple parameters, e.g. UV light security features may in some passport designs degrade naturally and thus their absence may not necessarily be a reason for suspicion.

Each company brought their own database (except Grabba which does not use a database). It could be surmised that some countries are in a position to develop and maintain better and more complete reference databases than private industry is able to, which will have an impact on the error-rate.

The impact of the software is also evident from the fact that two integrators who used the same scanner hardware but different software produced very divergent results mainly as a consequence of one managing to check less documents than the other in the given time (15 seconds per document).

The thresholds that are pre or manually configured and other calibration of the reader is also relevant, as one can locally at each border crossing point optimize the settings for the documents of the most common nationalities at the BCP. For example, certain passports of one nationality are very UV bright, and if the alert for UV brightness is set at the level required not to raise the alarm for these passports then presumably no UV alerts will ever be raised.

Another caveat is that in some cases industry reported a document as false since it was “expired”. The research team did not expect this, and thus had not made provisions for it. For those who listed “expired” as the only reason for failing a document we chose to interpret the answer as “genuine”. Nevertheless this aspect introduces a certain amount of noise in the results: for example it is not known if some of the readers have a decision tree where more detailed checks are halted once the first error is found (in this case the expiration date of the document), and how much (if at all) it affected the answers for the document reader (only one) which did not provide any detection points in the answers.

A fourth variable outside the reader system itself is the operator experience. Several of the operators (30%) had a border guard background, and this makes it difficult to decouple what in the results is due to the machine, and what is based on the experience of the operator.

In general it appears that the document readers were not as fast as expected from presentations of their capabilities, and in some cases they required human intervention in the user interface (selecting menu-items with the mouse, etc) before a conclusion could be drawn, further slowing down the process.

The machines easily identified manipulated or false MRZ. On the other hand, they generally failed to detect wrong printing techniques (for instance the forgery had laser-printing instead of engraving) and physical damage (e.g. damage to laminate safeguards).

3) Comparison Human-Machine Performance

The document experts made fewer mistakes, and processed a larger amount of passports in the given time, than the document readers. However, the results are very tentative and there are a number of uncertainties that need to be resolved in future studies before any firm conclusions can be drawn.
C. Methodological Lessons Learned

Based on the experience of this relatively simple and quick exercise, the following weaknesses should be corrected in follow-up research:

- Elaborate a formal metrics to evaluate performance and establish objective methods to rigorously collect all available data (example time needed, accuracy etc.);
- Collect reliable and consistent information on detection points either by allowing more time for reporting or directly interview the expert in the case of border control officer; by recording each decision and result from the automated inspection systems. The challenge here is to keep the examination from the reporting time as separate as possible in order to maintain the time pressure component;
- Involve a bigger number of first-line border control officers and thoroughly survey their background/skills/performance;
- Test the equipment and record results without the intervention of industry/expert operators;
- Compare the results with a full forensic analysis of the documents under examination;
- More genuine documents need to be included in the exercise to better reflect a more realistic ratio of genuine to false documents which is more in line with what first line officers would experience in real life.

VI. CONCLUSION

The exercise was instrumental in raising the awareness of both document experts and industry providers of automated document inspection systems of their respective strengths and weaknesses. Document experts who usually operate behind the lines were reminded of how difficult the decision-making in the first line is; how useful machines are in detecting abnormalities in security features not identifiable through human senses alone (optical or electronic); and how important updated training and specific education in understanding and using the machines are. Industry participants could use the occasion to demonstrate to practitioners how their equipments work, but also to better understand the importance of operational constraints such as time; the need to pay attention to physical security features, interfaces and training of the operators; and finally, from a more technical perspective, what major weaknesses should be remediated in order for their products to be market-proof.

Despite its original ambitious objectives, this exercise only managed to scratch the surface of the challenge of evaluating the performance of operational capabilities in order to shed some light on the detection dilemma. Future research needs to be more methodologically rigorous and study in depth all the variables that determine human and machine performance (errors), collect reliable operational data, and test different scenarios of human-machine cooperation in order to assess what combination proves to be the less vulnerable to attacks. Without a serious and well-researched consideration of the human factors involved in the task of document inspection in particular and border control in general, the use of automation may become a risk and reduce rather than increase border security.

REFERENCES