On the Vulnerability of Face Recognition Systems Towards Morphed Face Attacks

Ulrich Scherhag*, R. Raghavendra†, K. B. Raja†, M. Gomez-Barrero*, C. Rathgeb*, C. Busch*
* da/sec - Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany
Email: {ulrich.scherhag,marta.gomez-barrero,christian.rathgeb,christoph.busch}@h-da.de
† Norwegian Biometrics Laboratory, NTNU, Gjøvik, Norway
Email: {raghavendra.ramachandra,kiran.raja}@ntnu.no

***Abstract*—Morphed face images are artificially generated images, which blend the facial images of two or more different data subjects into one. The resulting morphed image resembles the constituent faces, both in visual and feature representation. If a morphed image is enroled as a probe in a biometric system, the data subjects contributing to the morphed image will be verified against the enroled probe. As a result of this infiltration, which is referred to as morphed face attack, the unambiguous assignment of data subjects is not warranted, i.e. the unique link between subject and probe is annulled.**

In this work, we investigate the vulnerability of biometric systems to such morphed face attacks by evaluating the techniques proposed to detect morphed face images. We create two new databases by printing and scanning digitally morphed images using two different types of scanners, a photo scanner and a line scanner. Further, the newly created databases are employed to study the vulnerability of state-of-the-art face recognition systems with a comprehensive evaluation.

I. Introduction

Biometrics based security solutions are widely deployed in various access control applications. In particular, face recognition represents a well-established and widely accepted method, since the reference is easy to capture: a standard camera can be used as a biometric sensor. Further, it can be verified by a human observer in a one to one comparison. One of the most prevalent use-cases for biometrics is automatic border crossing facilitated with biometric passports (ePass). The aim of the ePass is to strengthen the link between the physical document and its holder utilizing biometrics, which is envisioned to eventually reduce the risk of transfer or misuse of the document by a third person. In order to fulfill these demands, the ePass, which is designed according the recommendations of the International Civil Aviation Organization (ICAO), is capable of storing meta-data along with biometric information.

It was shown recently, that the issuing protocol of the ePass exhibits a security issue [1], [2], [3]. The key deficiency in the passport issuance process lies in the way the facial picture of an applicant is processed. In many countries, the applicant provides a printed facial image which is scanned and then digitally transferred to the passport production site (see Figure 2). As the facial image is provided by the applicant, it can be manipulated prior to the disposal at the federal offices. As a consequence, a specific attack scenario are morphed face attacks. An artificial facial image, which is referred to as morph, is created by blending the facial images of two or more different data subjects into one (see Figure 1). If the newly generated facial image is enroled to a Face Recognition System (FRS), the subjects contributing to the morphed image are positively verified against the morphed face attack reference [1], as the resulting morphed image resembles the constituent faces, both in visual and feature representation. Exploiting this fact, a black-listed subject (criminal) is able to obtain a legitimate ePass, by morphing his facial image with that of a non listed subject (accomplice), which the accomplice utilizes to apply for a passport. Due to the infiltration during the issuance process, the accomplice, as well as the criminal, are able to verify successfully against the reference stored in the ePass [1]. The feasibility of such morphed face attacks has been empirically confirmed in [3] and [1], [2].

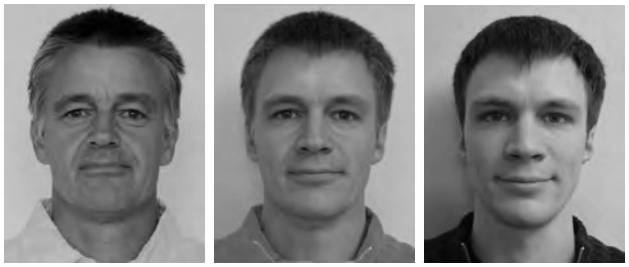

(a) Subject 1 (b) Morph Subj. 1 + 2 (c) Subject 2

Fig. 1: Example of face morphing

The vulnerability of a commercial FRS, with respect to morphed face attacks, was first explored in [1]. For the morphing process, landmarks in both faces are detected and moved towards each other. For a morphed face image, representing both faces in equal parts, the landmarks of the morphed face image are typically the mean of the features of both constituent original facial images. Among the many intermediate morphings that could in principle be used, the best morphed face between the two contributing images is further selected analyzing the comparison scores obtained from a FRS. Afterwards the skin and hair color have to be adapted, and some manual retouching can be done to remove artifacts. The experiment was carried out on a small database comprising 10 subjects with two commercial FRS [1], [2]. Another recent work in this direction has explored morphed face attack detection on digital images [3]. The threat of morphed face attacks was proven on

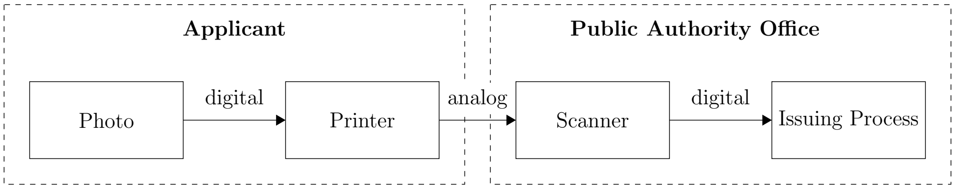


Fig. 2: Application process for the ePass

a database consisting of 450 morphed face images and twice as many normal face images, using a commercial FRS [3]. The morphing process was based on the method used in [1], with the difference that the mean face image was utilized instead of obtaining the morphed face image on the basis of comparison scores. In addition, a digital morphed face attack detection system was proposed employing Binarized Statistical Image Features (BSIF) [4], which represents a general purpose image descriptor, and a Support Vector Machine (SVM) with linear kernel as classifier.

The morphed face attack detection algorithm proposed in [3] focuses on morphed face attack samples in digital representation, such as possible for VISA application based on electronic photos or ePass renewal in New-Zealand. However, the application process of the ePass in many countries (e.g. most of the European Shengen states) still requires a printed face image that will be handed over to the public authority office during the application process, as depicted in Figure 2. A third deployed application procedure is to use live enrolment in the first place. In case of live-enrolment under supervision of an officer, the morphed face attack is not relevant. Unfortunately, only very few countries to date have decided to establish live-enrolment and many that have live-enrolment operate kiosks in an unsupervised manner. This procedure should be changed soon. However, in the meantime, the majority of countries continued to operate the hazardous second procedure, where face image is captured by a photographer or private person, printed and handed to the public authority office, where it is scanned again, a process compliant to the ICAO-standard [5]. Thus, morphed face attack detection is not limited to the digital domain, but extended to the detection of morphing attacks after printing and scanning of the morphed face attack sample, which we assertion to be more challenging, as during the processing of the face images information are lost and noise and granularity are added. To the best of our knowledge, this is the first work addressing the aforementioned real-world problem. The goal of this paper is thus to investigate whether the scientific findings on morphing and morphed face attack detection are transferable to the real-case scenarios. In particular, the following questions will be investigated:

- To which extent do morphed face attacks pose a threat to different (both commercial and academic) FRSs after printing and scanning?
- Does the type of the scanner (photo and line scan) influence the performance of the FRSs for morphed images?
- How effective are currently proposed morphed face attack detection algorithms [3] in detecting morphed face attacks after printing and scanning?

Answering these questions has led to the main contributions summarized here:

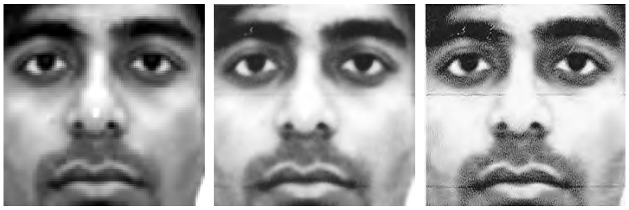
- A new database of morphed images is presented, which consists of printed and scanned images derived from digitally morphed images.
- Extensive evaluation of the vulnerability of two different FRSs with respect to scanned morph images is carried out.
- A comparative study on different currently proposed morphed face attack detection algorithms on the scanned morph face database is carried out to gauge the applicability and generalizability.

Further, in Section II, the newly constructed database of scanned morphed face images is presented. Section III shows the performance tests on FRSs and morphed face attack detection algorithms using the created database. Finally, in Section IV the key findings derived from this work are listed.

II. Databases

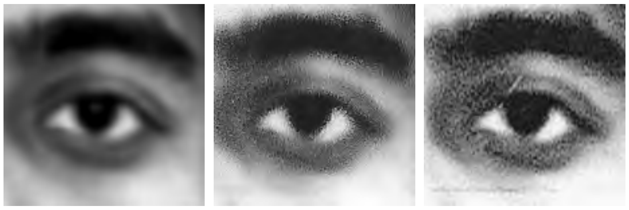
As the earlier works have focused on detecting morphed faces in raw format (digital), there exists no database to address the problem of detecting a morphed face attack after the print and scan process (see Figure 2). Thus, in this work, we have constructed a new database of morphed face images by first printing the digitally morphed images and scanning them afterwards with different devices. We started with the existing digitally morphed face database introduced in [3], consisting of cropped and grayscale images. The original database exhibits minor design issues, by correcting these issues the database comprises a total of 431 morphed images and 104 bona-fide (i.e. normal non-morphed) face images.

The database has been constructed in accordance with the guidelines of the International Civil Aviation Organization (ICAO) [5]. Specifically, the standard for the deployment of biometric identification in electronic Machine Readable Travel Documents (eMRTD) strongly advices the facial image to be scanned with at least 300 dpi and with a distance of approximately 90 pixels between the eyes. The guidelines also indicate the height of the face to be between 32 mm and 36 mm, which also complies with the ISO/IEC 19794-5 [6]. In order to create a realistic database, the aforementioned guidelines were followed to generate scanned facial pictures that met the requirements of ePass and ICAO photo specifications.



(a) Original (b) HP-Scan (c) RICOH-Scan

Fig. 3: Example of printed and scanned face images



(a) Original (b) HP-Scan (c) RICOH-Scan

Fig. 4: Close-up of the eye region of the original and scanned face images

Thus, the face images from the digital morph dataset are scaled to 32 mm and 36 mm. The images were then printed using a high quality 600 dpi ink-jet printer (*HP Photosmart 5520*). To simulate a realistic face image presented at a passport application, the printing was carried out on premium quality photo paper (HP premium plus photo paper). The printed face images were further scanned with two different types of scanner operating on two different principles, namely: i) a photo scanner integrated in the *HP Photosmart 5520*, and ii) a stand-alone scanner *RICOH MPC 6003 SP* with auto feed. While the images scanned with the *HP Photosmart 5520* scanner are henceforth referred to as HP-DB, the images acquired with the *RICOH MPC 6003 SP* are referred to as RICOH-DB. The total composition of the database is presented in Table I.

TABLE I: Composition of created database

Database	# Bona Fide Images	# Morphed Images
Digital DB	104	431
HP-DB	104	431
RICOH-DB	104	431

Figure 3 shows (a) the original face image and the scanned version of the same face (b,c). Even though high quality equipment was used for the process, a degradation of contrast and the addition of printing artifacts can be observed. The magnification of the right eye region in Figure 4 highlights the noise and granularity which is added by the printing and scanning process.

III. EXPERIMENTS

In this section, we describe the experimental protocol and the corresponding result of the evaluation on the newly constructed scanned morphed face image dataset. We first present the vulnerability of FRSs to scanned morphed face images and further show a comparative analysis of four different morphed face attack detection schemes [3] to detect scanned morphed images. We first list the metrics used to determine the error rates of FRSs when the scanned morphed face images are presented.

A. Performance evaluation metrics

The experiments are evaluated according to the metrics introduced in ISO/IEC 19795-1 [7], ISO/IEC 2382-37 [8] and ISO/IEC FDIS 30107-3 [9], which are summarized here.

The vulnerability of different FRS in this work will be reported using:

Impostor Attack Presentation Match Rate (IAPMR): in a full-system evaluation of a verification system, the proportion of impostor attack presentations using the same Presentation Attack Instrument (PAI) species in which the target reference is matched [9].

The performance of the detection systems will be analyzed with a Detection Error Trade-off (DET) curve reporting:

Attack Presentation Classification Error Rate (APCER): proportion of attack presentations using the same PAI species incorrectly classified as bona fide presentations in a specific scenario [9].

Bona Fide Presentation Classification Error Rate (BPCER): proportion of bona fide presentations incorrectly classified as presentation attacks in a specific scenario [9].

In addition we will report:

BPCER10: proportion of bona fide presentations incorrectly classified as presentation attacks for a fixed APCER of 10% [9].

BPCER20: proportion of bona fide presentations incorrectly classified as presentation attacks for a fixed APCER of 5% [9].

Further used metrics will be:

False Acceptance Rate (FAR): error of accepting a biometric claim that should have been rejected in accordance with an authorative statement on the origin of the biometric probe and the biometric reference [7].

False Reject Rate (FRR): proportion of verification transactions with truthful claims of identity that are incorrectly denied [7].

B. Vulnerabilities of Face Recognition Systems

In this study, we consider two different FRSs to demonstrate the vulnerability and deficiencies towards detecting scanned morphed face images using the newly constructed database. For our experiments we use both a commercial system - Neurotechnology VeriLook SDK [10] - and a publicly available FRS - OpenFace [11] - based on a pretrained Deep Neural Network (DNN). In order to allow a fair comparison

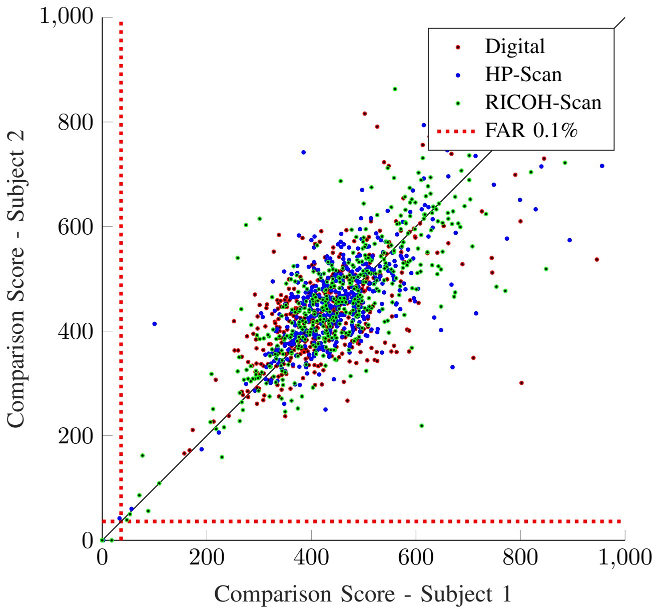


Fig. 5: Scores obtained by comparing morphed images to each of their contributing original images with VeriLook

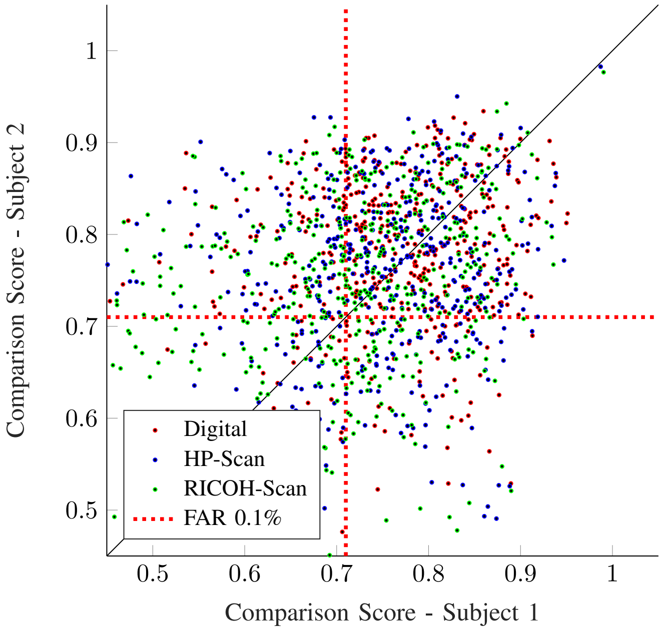


Fig. 6: Scores obtained by comparing morphed images to each of their contributing original images with OpenFace

with earlier works [1], [3], the vulnerability analysis on FRS is depicted in Figure 5 and 6 with scatter-plots. Each data point represents the result of a mate comparison trial (i.e. genuine comparison) involving a morphed image and each of the subjects contributing to the morphing. In other words, each coordinate is determined by the score obtained from comparing the morphed image against a probe image of one of the two constituent subjects. The configuration of the FRS in this work follows the guidelines of European Agency for the Management of Operation Cooperation at the External Borders of the Member States of the European Union (FRONTEX), that recommends the operation point of FRS in border control scenarios at a fixed FAR of 0.1%.

In both figure, the similarity scores depicted are referred to as Digital (in red), and both scanned databases, referred to as HP-Scan (in blue) and RICOH-Scan (in green), the read dashed lines represent the verification threshold at a FAR of 0.1%.

Figure 5 depicts the similarity scores generated by the VeriLook algorithm on the database presented in [3]. Based on the recommendations from Neurotechnology, a threshold of comparison score equal to 36 (corresponding to FAR of 0.1%) was employed. It can be observed, that all images from the Digital-DB achieve scores far above the verification threshold, thereby resulting in a positive match. In particular, the lowest score of the complete dataset equals 209, in comparison with the verification threshold set at 36.

Figure 6 depicts the similarity scores generated by the OpenFace algorithm. There is no threshold specified for the FAR 0.1% for this database, so a threshold for FAR 0.1% was determined using a disjoint dataset - FRGCv2.0 database [12], which comprises 457 subjects with a total of 25 091 images. The OpenFace algorithm returns distance scores, which were normalized to similarity scores by calculating the inverse Min-Max normalization where the minimum and the maximum comparison score values are determined on the FRGCv2.0 database [12]. Employing the threshold derived using OpenFace on FRGCv2.0 database [12] at FAR of 0.1% has resulted in FRR of 10.81%.

TABLE II: IAPMR of Face Recognition Systems

Algorithm	**Digital-DB**	**HP-DB**	**RICOH-DB**
VeriFace SDK	100%	99.7%	97.3%
OpenFace	80.1%	62.3%	70.8%

As it may be observed, after the printing and scanning process, some of the comparison scores lie below the verification threshold. However, the vast majority of them would still grant a positive verification of the constituent subjects, thus proving the vulnerability of the analysed FRS to scanned morphed images. In fact, if the FRS is evaluated regarding its PAD capabilities, the IAPMR can be considered as an indication of the *attack success* chances. Table II presents the IAPMR for the images in different subsets of the newly created database. It can be noted that 100% of the images from the Digital-DB are successfully compared while there is 99.7% success for images from HP-DB and 97.3% success for images form RICOH-DB using VeriLook SDK. Lower values of IAPMR can be seen when OpenFace is employed on the three data subsets. As a note, a relatively lower performance in terms of IAPMR of scanned images can be attributed to the artifacts resulting from the printing and scanning process, but it is likely that a genuine comparison of images exhibiting the same scan artifacts might be rejected as well, as the baseline performance of OpenFace is far below the baseline of VeriLook SDK.

C. *Performance of Morphed Face Attack Detection Algorithms*

As shown in Section III-B, state-of-the-art face recognition algorithms are highly vulnerable to morphed face attacks. Thus, it is essential to detect such attacks and tackle them to improve the reliability of FRS. Based on the very high classification rate reported earlier to detect morphed images in digital format [3], we evaluate three different morphed face attack detection schemes. In order to have a fair benchmark, we employ 200 images for training the SVM classifier and rest of the 231 images for testing the accuracy of the different algorithms. It should be noted that both sets of morphed images are disjoint. The results obtained using different schemes are presented in Table III, which already shows a weak detection accuracy on the new Digital-DB with *BPCER10* of 46.1%.

The same algorithms were checked against the generated HP-DB and RICOH-DB. It can be noted from Table III that various algorithms perform poorly on detecting the morph images after scanning (*BPCER10* between 37% and 89% for the HP-DB and *BPCER10* between 40% and 65% for the RICOH-DB). The lower performance highlights the challenging nature of detecting morphed face images with local image descriptors. For the sake of better visualization, the Detection Error Trade-off (DET) plot of the best performing algorithm (BSIF-SVM) is depicted in Figure 7. We may observe that the algorithm shows similar performance on both scanners (in blue and green, with Equal Error Rates (EER) of 22.2% and 20.8%) and on the Digital-DB (red, EER = 24.2%).

IV. Conclusion

Printed and scanned morphed face attack images are still a major threat to FRSs. Currently proposed morphed face attack detection algorithms are neither successful on digital morphed images, nor when the attack samples are printed and scanned. This work has thus demonstrated the difficulties in detecting morphed face images after print and scan with a thorough experimental validation.

On the basis of the experiments, the research questions posed in the introduction can be answered as follows:

1) Printing and scanning of the morphed face images has only a slight effect on the IAPMR of FRS. The systems are still highly vulnerable to the morphed face attacks.
2) Irrespective of the type of scanner employed (photo or line-scanner), morphed images can easily succeed in attacking FRS. For the Commercial Off-The-shelf (COTS) system, the IAPMR was lowered by 0.3% absolute for the photo scanner and 2.7% absolute for the line-scanner. For the open-source face recognition system, the IAPMR was reduced by 17.8% and 9.3% absolute, still posing a major threat to FMR systems.
3) Currently proposed morphed face attack detection systems are neither successful in detecting morphed face images in digital format, nor when they are printed and scanned. For both scenarios the *BPCER10* is far above 30%.

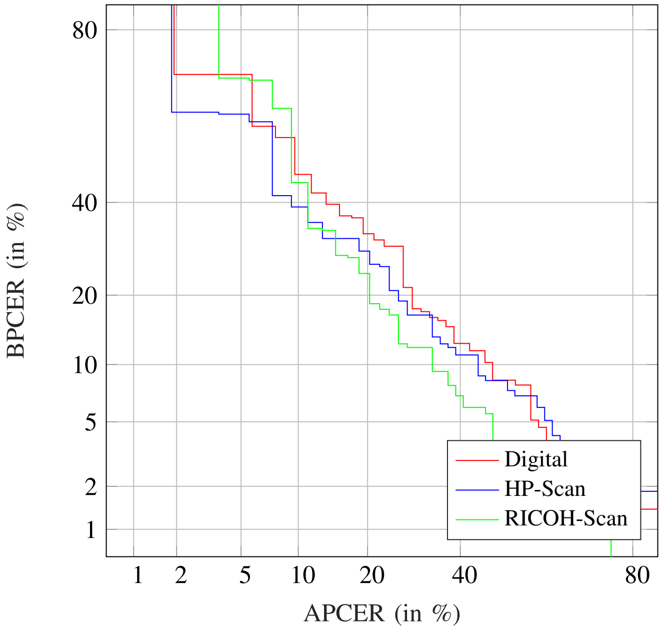


Fig. 7: DET-Curve of detection algorithm (BSIF-SVM) [3]

In order to develop morphed face attack detection algorithms, which are able to achieve a stable performance on printed and scanned images, further research is necessary. Also the impact of printing and scanning of facial images to FRS should be analyzed in this regard. Since in this work a fixed resolution for printing was used, further research on different scanning resolutions may reveal new perceptions.

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within CRISP. Raghavendra Ramachandra and Kiran B Raja are supported by Research Council of Norway (Grant No. IKTPLUSS 248030/O70).

TABLE III: BPCER of morphed face attack detection at fixed APCER

		BPCER10			BPCER20	
Metric	**Digital DB**	**HP-DB**	**RICOH-DB**	**Digital DB**	**HP-DB**	**RICOH-DB**
BSIF-SVM [3]	46.1%	36.9%	39.3%	61.5%	50.6%	63.8%
LBP-SVM [3]	57.9%	76.1%	57.6%	68.3%	86.0%	69.3%
LPQ-SVM [3]	68.3%	88.8%	65.0%	76.4%	94.4%	73.7%

References

[1] M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *IEEE International Joint Conference on Biometrics*. IEEE, sep 2014, pp. 1–7.

[2] ——, "On the effects of image alterations on face recognition accuracy," in *Face Recognition Across the Imaging Spectrum*. Springer Nature, 2016, pp. 195–222.

[3] R. Raghavendra, K. B. Raja, and C. Busch, "Detecting Morphed Face Images," in *8th IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2016.

[4] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," *21st International Conference on Pattern Recognition (ICPR)*, no. Icpr, pp. 1363–1366, 2012.

[5] International Civil Aviation Organisation, "Machine Readable Travel Documents - Part 9: Deployment of Biometric Identification and Electronic Storage of Data in eMRTDs," ICAO, Montreal, Tech. Rep., 2006.

[6] International Organization for Standardization, "Information Technology – Biometrics – Biometric Data Interchange Formats – Face Image Data," JTC 1 /SC 37, Geneva, Switzerland, Tech. Rep. 19794-5, 2005.

[7] ——, "Information technology – Biometric performance testing and reporting – Part 1: Principles and framework," JTC 1/SC 37, Geneva, Switzerland, ISO/IEC 19795-1:2006, 2006.

[8] ——, "Information technology – Vocabulary – Part 37: Biometrics," JTC 1/SC 37, Geneva, Switzerland, ISO/IEC 2382-37:2012, 2012.

[9] ——, "Information Technology – Biometric presentation attack detection – Part 3: Testing and reporting," JTC 1/SC 37, Geneva, Switzerland, ISO/IEC FDIS 30107-3:2017, 2017.

[10] Neurotechnology, "VeriLook SDK." [Online]. Available: http://www.neurotechnology.com/verilook.html

[11] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.

[12] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, vol. I, pp. 947–954, 2005.