Humans Vs. Algorithms: Assessment of Security Risks Posed by Facial Morphing to Identity Verification at Border Control

Andrey Makrushin, Tom Neubert and Jana Dittmann
Otto-von-Guericke University of Magdeburg, Universitaetsplatz 2, Magdeburg, Germany

Keywords: Facial Morphing, Automated Border Control, Human Experiment, Face Recognition, Morphing Detection.

Abstract: Facial morphing, if applied to a biometric portrait intended for an identity document application, compromises further identity verification by means of the issued document. An electronic machine readable travel document is a prime target of a face morphing attack because a successful attack allows a wanted criminal for illicit border crossing. The open question is whether human examiners and algorithms can be fooled only by professionally created manual morphs or even by automatically generated morphs with evident visual artifacts. In this paper, we introduce a border control simulation to examine the ability of humans in recognizing morphed passport photographs as well as in mismatching morphed passport photographs against "live" faces of travelers. The error rates of humans are compared with those of algorithms to emphasize the necessity for computer-aided support of border guards.

1 INTRODUCTION

The traveler identity verification done by border guards is slow and prone to errors (White et al., 2014). The recent trend to speed up border crossing and save expensive manpower is the deployment of Automated Border Control (ABC) systems also known as electronic gates. Modern ABC systems rely on two-factor identity verification including an authenticity and integrity check of an electronic Machine Readable Travel Document (eMRTD) and biometric authentication of a traveler. As early as 2002 the International Civil Aviation Organization (ICAO) selected face to be the primary biometric trait used with eMRTD (ICAO, 2004). Technically, an automated face recognition (AFR) system that is integrated in electronic gates captures a "live" face and compares it with a digital passport photograph stored on a chip of an eMRTD.

An automation of border control attracted the interest of security experts due to the risk that wanted criminals practice presentation or morphing attacks for illicit border crossing. While the biometric research community has for a long while been concerned with the face presentation attack (Raghavendra and Busch, 2017), the face morphing attack is a novel and more sophisticated fraud whose potential harm significantly exceeds that of the face presentation attack (Kraetzer et al., 2017).

The face morphing attack includes two steps. Let us assume that Mallory is an attacker and Allice is her accomplice. First, Alice applies for a new identity document with a morphed face image of her and Mallory's faces. If an officer accepts the image, the issued document is authentic and perfectly regular. Second, Mallory uses the document for identity verification claiming to be Alice.

A morphed face image is a key to a successful morphing attack. It is shown in (Ferrara et al., 2016) that manually created morphs preserve facial characteristics of all contributing faces so that both humans and algorithms often falsely match a morphed image against genuine images of any of contributing faces. Visually faultless facial morphs can be generated even automatically, and humans are almost unable to recognize these as such (Makrushin et al., 2017).

It is clear that high-quality morphs pose a threat to the identity verification process no matter whether it is completely automated or maintained by human examiners. The open question is whether automatically generated morphs with visual artifacts bear the risk of being accepted by humans and algorithms. For instance, the study in (Robertson et al., 2017) demonstrates that the error rates of humans with automatically generated low-quality morphs are not as dramatic as in (Ferrara et al., 2016) with manually generated high-quality morphs.

Makrushin, A., Neubert, T. and Dittmann, J.
Humans Vs. Algorithms: Assessment of Security Risks Posed by Facial Morphing to Identity Verification at Border Control.
DOI: 10.5220/0007378905130520
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 513-520
ISBN: 978-989-758-354-4
Copyright © 2019 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

Aiming at answering this question we created an experiment that simulates a border control scenario. First, we let people perform *morphing detection* namely to judge whether a face image is morphed or genuine and compare the results with those of forensic morphing detectors. Second, we let people perform *biometric matching* of a potentially morphed passport photograph with a "live" face and compare the results with those of AFR systems.

Our main contribution is an implementation and deployment of the ongoing web-based experiment which attracted more than 400 participants within the first two days after start. The evaluation of the exp. results allows us for the following conclusions:

- biometric matching is significantly more prone to errors than morphing detection;
- skilled participants perform on average slightly better than unskilled;
- algorithms demonstrate lower error rates than humans in both tasks.

Hereafter, we summarize papers addressing human experiments with facial morphs in Section 2. In Section 3, we describe our border control simulation in detail. Section 4 comprises description of AFR systems and morphing detectors used as a reference. In Section 5, we compare error rates of humans and algorithms. Section 6 concludes the paper with the summary of results.

2 STATE OF THE ART

Considering a questioned passport photograph, there are two tasks that arise: (i) matching against a person who presents the photograph and (ii) morphing detection. The former one (i) is a well-studied task from the field of biometrics for which the recognition performances of humans and AFR systems have been compared in different scenarios (Phillips and O'Toole 2014) including border control (del Rio et al., 2016). The latter task (ii) is new so that there is a lack of studies comparing morphing detection performances of humans and automated detectors.

Morphing detection could be "blind", meaning it is based solely on the presented photograph, or could rely on a reference face image. Note that in almost all scenarios, it is possible to take a reference face image. However, many detection algorithms ignore this option. Technically, the setups for biometric matching and morphing detection with a reference does not differ. Both processes operate on two images (a document image and a "live" image) and as a result either accept or reject a person. The only difference is the reason of rejection - *no match* or *morphed.* Since a border guard solves both tasks simultaneously, these can be fused in an experiment.

The first study on comparison of humans and algorithms to perform biometric matching with morphed face images is conducted in (Ferrara et al., 2016). The authors generated 80 morphs and asked 44 border guards and 543 laymen to match those against original faces. Surprisingly, border guards did not perform better than laymen. They accepted on average slightly more genuine trials (91.67% vs. 87.76%) but also significantly more morphing trials (74.92% vs. 57.55%). All in all, both border guards and laymen have demonstrated unacceptably high morph acceptance rates (MAR). Three commercial AFR systems were examined with the same images. The MAR values were dramatically high, revealing the complete inability of the systems to reject morphed faces. Later on, high MAR of two AFR systems were confirmed in (Scherhag et. al, 2017). However, we believe that this estimation of MAR is pessimistic because the images of a person used for morphing and matching are very similar (no variance regarding pose and illumination). The images are also not as rich in detail as biometric face images intended for documents, making morphing artifacts (e.g. ghosting) visually less perceptible.

A realistic experiment would require "live" images with random background and illumination for matching. Such experiments have been conducted in (Robertson et al., 2017) and (Robertson et al., 2018). In the former study, test participants first matched face images having two options: accept or reject a verification trial mixing up biometric matching and morphing detection with a reference, and then having three options: accept, reject because of no match, and reject because the passport image is morphed. The MAR dropped from 68% in the first experiment to 21% in the second. This reveals the fact that if examiners are aware of potential morphing in passport photographs, the acceptance of impostor trials is less probable. In the latter study, the authors investigate how much human error rates drop after coaching.

The study in (Makrushin et al., 2017) reports the results of the first human experiment on blind morphing detection. The participants should detect morphing in photographs printed with a passport dimension of 35x45 mm. The resulting average MAR was 44.6% and the FRR 43.64% which is not far from random guessing. The high FRR can be explained by reluctance of participants to skip morphed images.

In general, all aforementioned studies manifest the necessity for automated morphing detectors to support staff at document issuing offices to prevent issuing of double-identity documents as well as at document checking stations to withdraw double-identity documents from circulation. Nonetheless, the development of dedicated morphing detectors is in its early phase, but researchers have already made remarkable progress in designing prototypes. An overview of recently introduced face morphing detectors is given in (Makrushin and Wolf, 2018).

Although some human experiments have been conducted, it is still unclear how successful human examiners can be in detecting facial morphs and matching unfamiliar faces having in mind that one of both could have been morphed and whether algorithms perform worse or better. We try to fill this gap with our experiment.

3 THE BORDER CONTROL EXPERIMENT

Since we want to know how easy is it on average to deceive human examiners with a morphed face image, the main objective of our experiment is a realistic simulation of a border control from the viewpoint of a border guard. Our experiment is implemented as a web-based questionnaire and is available online at: https://bit.ly/2JdgvII. A smart-phone, a tablet or a regular computer could be used equally well to complete it. Going online allowed us to reach a large number of participants.

Assuming that border guards are trained to recognize the morphing attack, participants of the experiment should have at least basic knowledge of facial morphing. Therefore, the experiment starts with the brief explanation of the morphing process followed by a tutorial on how to detect morphed face images. Figure 1 shows two face images with typical morphing artifacts presented during the tutorial. The first one demonstrates the ghosting artifacts in the hair, at the temples, on the clothes and in the eyes/ irises, and the second one a "swimming cap" effect - an edge on a forehead resulting from the "non-optimal" splicing of a morphed face into the original background as well as ghosting artifacts in the eyes.

The questionnaire is divided into two parts, 15 questions each. In the first one (see Section 3.1), an examiner is asked whether the face on a passport photograph is morphed or not. In the second one (see Section 3.2), a passport photograph is presented to an examiner together with a video of a traveler approaching the passport check desk and the examiner is asked to match a person on the passport photograph against a person shown in the video.

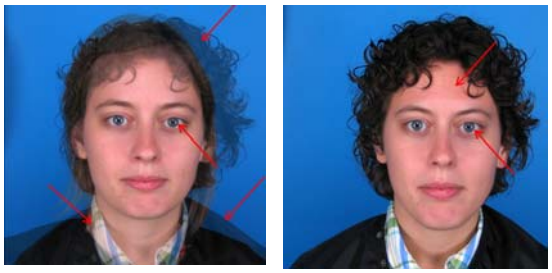


Figure 1: Examples of morphing artifacts: spurious shadows (ghosting) in hair, eyes and clothes regions, apparent transition between brow ridges and a forehead; Orig.images from http://pics.stir.ac.uk/2D_face_sets.htm.

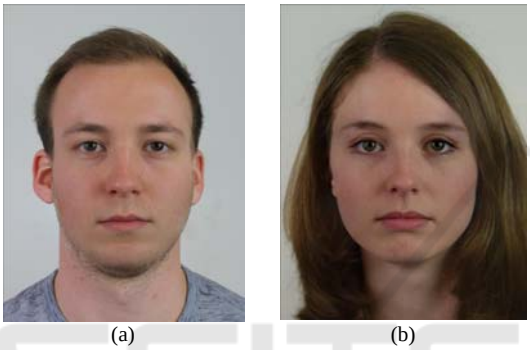
The passport photographs are compliant with the Portrait Quality Standard for reference facial images for MRTD maintained by ICAO (ISO/IEC JTC1 SC17 WG3, 2018). This means that a face is in frontal position, in-plate rotation angle does not exceed 5%, facial expression is neutral, face is in the middle of the image, the face size is in the certain proportion to the image size, and the illumination is uniform. In the first part of the experiment, we use high-resolution raw images and, in the second, the images scaled to 531x413 pixels to simulate photographs stored on the chip of an MRTD. The morphed face images used as passport photographs are generated automatically using the approaches from (Makrushin et al., 2017) and (Neubert et al., 2018). The morphed images have not undergone any retouching or post-processing and, therefore, may include apparent visual artifacts. We deliberately include morphs of different quality.

We split the human examiners to the groups of skilled and unskilled ones according to whether they are familiar with the morphing issue and compare the group performances. After the test is finished, we asked the examiners which face regions contributed the most to the decision. It helps us to better understand the human intuition about abnormalities in a face. To avoid biased decisions, we filter out examiners who know one or more donors of the photographs. We store neither personal data nor meta data of test participants except for the time required for each decision.

One month after launching, the final number of test participants exceeded 450. However, only 282 examiners know no one of the photograph donors. Among them there are 49 skilled, 230 unskilled and 3 provided no information on their experience.

3.1 Detection Experiment

Here, we challenge the humans to "blindly" detect morphed face images. An examiner looks at a passport photograph and decides whether it is morphed without any further knowledge about a person. In a real-life identity verification process, an examiner can make use of a meta data of the person. However, our idea is to avoid any biases and let people decide solely based on visual morphing artifacts. Two morphed images from the experiment are shown in Figure 2. The fist one (a), representing a transgender face, was correctly detected by approx. 97% of examiners and the second (b), representing a female face, by approx. 67%.



(a) (b)

Figure 2: Morphing detection in a passport photograph: (a) an "easy" morph M9 and (b) a challenging morph M2.

The 15 questions in the first part include 5 genuine (G1, G2, ..., G5) and 10 morphed face images (M1, M2,..., M10). The image resolution spread from 2.8 to 6.4 megapixels with one exception of 531x413 pixels. If human examiners would constantly achieve high recognition performance in this or a similar experiment, it could lead to a requirement to store high-resolution digital images on a chip of an MRTD. This part of the experiment can be seen as a kind of training for the next part - identity verification with potentially morphed photographs.

3.2 Matching Experiment

Here, we simulate a border control scenario. The humans are challenged to match passport images against "live" faces having in mind that the passport image could have been morphed. An examiner looks at a passport photograph and watches the video in which a traveler approaches the passport check desk and decides whether the person should be accepted or rejected. The decisions here are biased by the first part of the experiment because obvious morphing artifacts in a passport image would lead to rejection without comparing the faces. Two samples from the experiment are shown in Figure 3. The first one (a) was correctly rejected by approx. 72% of examiners and the second (b) by approx. 55%.

(a)

(b)

Figure 3: Matching of a potentially morphed passport photograph against a "live" face: (a) a less challenging sample and (b) a more challenging sample.

The 15 questions in the second part include matching trials with 6 genuine (G1, G2, ..., G6) and 9 morphed passport photographs (M1, M2, ..., M9). The resolution of all passport photographs is 531x413 pixels. The video resolution is 320x240 pixels. There was no special reason to select such a small resolution and we believe that the experiment could benefit from videos of higher resolution. The poor matching performance of human examiners in this or a similar experiment should discomfort the authorities responsible for security issues at border control and motivate them to make use of AFR systems and dedicated morphing detectors.

4 ALGORITHMS

In the following, we describe two established AFR systems and two recently introduced morphing detectors whose error rates will be compared with error rates of human examiners.

4.1 AFR Systems

As proponents of AFR systems, we selected one commercial off-the-shelf (COTS) solution - Luxand FaceSDK (https://www.luxand.com/facesdk/) and a face recognition tool provided in the Dlib - an open source programming library (http://dlib.net/).

We use the ***Luxand FaceSDK 6.5.1*** released in June 2018. For two incoming face images, the outcome of the matcher is a score in the interval from 0 to 1. Higher scores are for more similar faces. Based on the proprietary experiments, the SDK provides decision thresholds at which the False Acceptance Rate (FAR) does not exceed 1% and 0.1%. These are 0.99 and 0.999 respectively. Up to the version 6.5, the maintainer reports the True Acceptance Rate (TAR) of 99.85% at the FAR of 0.1% according to NIST FRGC testing (Luxand, 2018). Note that the FAR of less than 0.1% is recommended for ABC in (FRONTEX, 2015).

Starting from the release 19.3 in February 2017, the ***Dlib*** library includes a ***face recognition*** tool (King, 2018). The face classification model is built upon the slightly modified ResNet-34 network having 29 convolutional layers with half the number of filters in a layer (dlib_face_recognition_resnet_model_v1). The network is trained from scratch using about three million images gathered from Face Scrub and VGG datasets as well as some images from internet. Prior to feeding into the network, the face images are geometrically aligned and scaled to 150x150 pixels. The network maps images to a 128-dim vector space in which all identities are supposed to be represented by non-overlapping balls of radius 0.6. The images of the same identity should be close to each other and images of different identities far apart. The distances d returned by the matcher are in the interval from 0 to 1. With the distance threshold of 0.6, the model obtains an accuracy of 99.38% on the Labeled Faces in the Wild benchmark, which is as good as other state-of-the-art face recognition methods as of beginning 2017. We replace distances by similarity scores $s=1-d$ to make the matchers comparable. The decision threshold becomes 0.4.

4.2 Morphing Detectors

As the development of dedicated morphing detectors is in its early phase, there are currently no mature solutions on the market. However, researchers have already made remarkable progress in designing prototypes. Since the morphing detectors based on Deep Convolutional Neural Networks (DCNN) are confirmed to perform the best (Raghavendra et al., 2017), we examine a DCNN-based detector from (Seibold et al., 2018) and compare it with the keypoint-based detector from (Kraetzer et al., 2017).

The ***keypoint-based morphing detector*** relies on the assertion that the blending operation, which is an indispensable part of the morphing process, causes a reduction of face details. Hence, the number of significant corners and edge pixels is expected to become lower in morphed images in comparison to genuine ones. The detector comprises five keypoint detectors and two edge detectors:

- Scale Invariant Feature Transform (SIFT);
- Speed Up Robust Feature (SURF);
- Features from Accelerated Segment Test (FAST);
- Oriented FAST and rotated Binary Robust Independent Elementary Features (ORB);
- Adaptive and Generic Accelerated Segment Test (AGAST);
- Canny edge detector;
- Sobel edge detector for horizontal and vertical edges.

A feature is a number of keypoints/edge pixels detected in the face region that is a convex hull of the 68 facial landmarks extracted from the image by the *Dlib* shape predictor. Each feature is normalized by the natural logarithm of the number of pixels in the face region. This step is essential because the number of detected keypoints non-linearly increases with a face size. The normalization makes features invariant to image scaling. The 8 aforementioned features are extracted from an original image and from the same image after JPEG compression with the quality factor 0.75. The idea behind this is that for genuine images the compression leads to significant loss of details and for morphed images does not. The last set of 8 features comprises the ratios of the features in compressed and non-compressed images. Hence, an image is represented by a 24-dim feature vector. The linear support vector machine is trained based on a proprietary dataset of 2000 genuine and 2000 morphed high-resolution passport images. Facial morphs are created using approaches from (Makrushin et al., 2017) and (Neubert et al., 2018).

The ***DCNN-based morphing detector*** considered here is referred to as "naive" in the original paper. It is built upon the VGG19 network originally trained to classify images within the ILSVRC challenge and modified to a binary classifier by applying transfer learning. The training dataset includes approx. 1900 face images of different individuals gathered from several public databases and from the internet. Morphed face images were created from pairs of faces using two different approaches from (Seibold et al., 2017) taking into account that images are from the same database, individuals have the same gender and each image was used with equal frequency. The set of training images was augmented by the filtered versions of images applying the following filters: Motion blur, Gaussian blur, Salt-and-pepper noise,

and Gaussian noise. The numbers of genuine and morphed images in the training set are equal. Prior to feeding a face image into the network, it is rotated, such that the eyes are on the horizontal line, cropped to keep the region between eye brows and mouth and between the outer pairs of the eyes only, and scaled to 224×224 pixels.

5 EVALUATION

By August 6th, 2018, 9:00 a.m., the number of test participants was 477 resulting in 477 test protocols. We take this snapshot as a basis for the evaluation. Only 282 participants out of 477 know no one of the photograph donors, reducing the number of further processed test protocols. This filtering is important because any degree of familiarity between a border guard and a traveler is very unlikely and would lead to unwelcome bias in decisions. Out of 282 unbiased participants, 49 claimed to have some knowledge on face morphing (we call these participants skilled examiners), 230 claimed to encounter face morphing for the first time (we call these participants unskilled examiners), and 3 participants did not disclose their experience. Matching takes on average a little bit longer than detection - 16.7 vs. 13.4 seconds per sample. Skilled examiners are slightly faster than unskilled in detecting morphs and there is no clear trend who is faster in matching faces.

5.1 Biometric Matching Performance

The metrics for matching performance are adopted from biometrics. These are False Reject Rate/True Accept Rate (FRR/TAR) for genuine and Morph Accept Rate/True Reject Rate (MAR/TRR) for morphing trials. For AFR systems we also use Equal Error Rate (EER).

The matching rates of ***human examiners*** are introduced in Figure 4. On average, all examiners correctly rejected 65.35% of morphing trials. The skilled examiners with the average TRR of 67.57% performed only slightly better than unskilled examiners with the average TRR of 64.88%. The matching rates for genuine trials are noticeably better. On average, all human examiners correctly accepted 77.30% of genuine trials. The difference in matching rates between skilled examiners (average TAR of 78.23%) and unskilled examiners (average TAR of 77.10%) is negligible. The surprisingly low TAR resulting from our experiment is not far from those reported in the study on recognition of unfamiliar faces (Hancock et al., 2000) confirming once again that people are bad at matching unfamiliar faces.

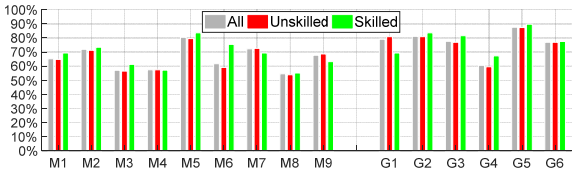


Figure 4: Matching rates of skilled (green), unskilled (red) and all (grey) human examiners; TRR for morphing trials M1-M9 and TAR for genuine trials G1-G6.

The error rates of ***Luxand FaceSDK*** and ***Dlib face recognition*** are shown in Figure 5. The EER of former one is 0% at the decision threshold of 0.9335 which means that the system would make correct decisions in all 15 trials. However, with the recommended threshold of 0.999, the system correctly rejects all morphing trials, but also falsely rejects 3 out of 6 genuine trials. The EER latter one is approx. 11.11% at the decision threshold of 0.4927. With the recommended threshold of 0.4, the system correctly accepts all genuine trials, but also falsely accepts 6 out of 9 morphing trials. At thresholds from 0.4824 to 0.4926, the detector makes no false rejections and falsely accepts only one morphing trial.

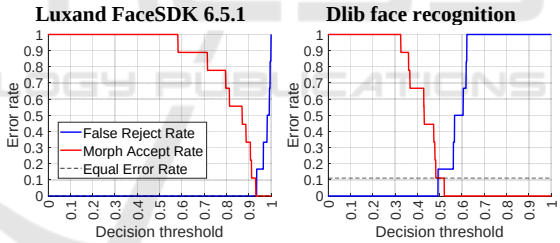


Figure 5: FRR/MAR curves of AFR systems.

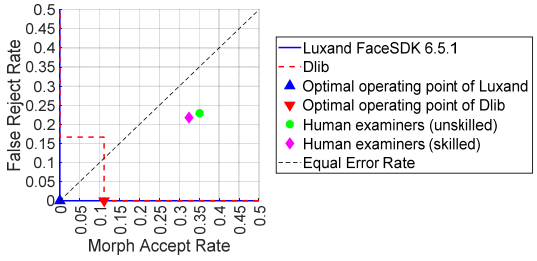


Figure 6: DET curves of Luxand FaceSDK and Dlib face recognition along with error rates of human examiners.

The DET diagram in Figure 6 shows the difference in matching rates of AFR systems and human examiners. Luxand FaceSDK has better matching performance than Dlib face recognition. With the

properly set decision thresholds both AFR systems significantly outperform human examiners. However, with the standard decision thresholds the matching performance of AFR systems is on the same level or even worse than that of humans.

5.2 Morphing Detection Performance

Since morphing detection is a binary classification task, the standard metrics are used: True Positive Rate (TPR)/False Negative Rate (FNR) for morphed images, and True Negative Rate (TNR)/False Positive Rate (FPR) for genuine images. Note that morphed images are seen as entities of the positive class. We also use Half Total Error Rate (HTER) to compare performances in an easier way.

The matching rates of ***human examiners*** are shown in Figure 7. On average, all human examiners correctly detected 84.59% morphed images. The skilled examiners with the average TPR of 88.78% performed noticeably better than unskilled with the average TPR of 83.70%. The detection rate also strongly depends on the quality of a morphed image. For instance, the "easy" morph M9 in Figure 2a was correctly detected by 89.57% unskilled and by 95.92% skilled examiners, while the challenging morph M2 in Figure 2b by only 67.39% unskilled and 67.35% skilled examiners. Note that for the challenging morph, the detection rates of skilled and unskilled examiners do not differ significantly. Based on this result, we conclude that special training could improve the human ability to detect morphs, but cannot be seen as a panacea, since for many high quality morphs the detection rates of skilled and unskilled examiners are similar. For genuine face images, the average detection rates of skilled and unskilled examiners are very close to each other yielding 83.27% and 82.35% TNR respectively. As morphing indicators, successful examiners most frequently selected artifacts on eyes followed by artifacts in a forehead/temple region, and only few examiners pointed to artifacts in hair, clothes and background.

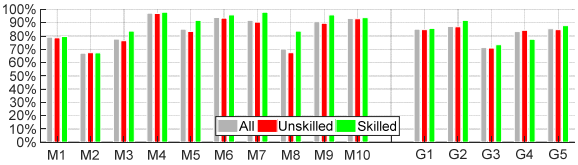


Figure 7: Detection rates of skilled (green), unskilled (red) and all (grey) human examiners; TPR for morphing trials M1-M10 and TNR for genuine trials G1-G5.

The estimate for the EER of the ***keypoint-based detector*** is 10% obtained with the decision threshold of 0.9735 (see Figure 8). In our test, however, using this threshold leads to 9 (out of 10) correctly detected morphs, and 1 (out of 5) false alarm for genuine images. The minimal Half Total Error Rate (HTER) of 5% can be achieved with a threshold between 0.9739 and 0.9790 meaning no errors for genuine images (100% TNR) and 9 (out of 10) correct decisions for morphs (90% TPR). With the recommended decision threshold of 0.5, the detector yields 100% TPR, but only 20% TNR.

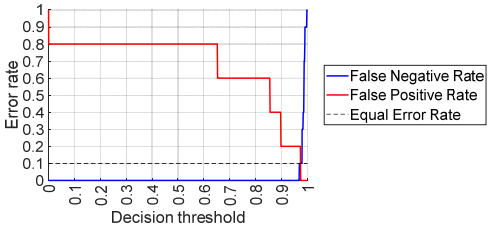


Figure 8: Error rate curves of the Keypoint-based detector.

The ***DCNN-based detector*** produces scores in the interval [0; 0.000000015] for the genuine images and in the interval [0.9997; 1] for the morphed images, allowing for a clear separation with almost any threshold deviating from 0 and 1. With the recommended threshold of 0.5, the detector makes literally no mistakes and can be seen as a very robust tool for morphing detection.

The DET diagram in Figure 9 demonstrates the difference in error rates between the two considered morphing detectors and the averages of skilled and unskilled human examiners. Considering HTER, the performance of skilled examiners is approx. 14% and of unskilled examiners approx. 17%. The HTER of keypoint-based detector at the decision threshold of 0.975 is 5%. The DCNN-based morphing detector perfectly solves the problem yielding the HTER of 0%. Hence, with the properly set decision thresholds, both morphing detectors significantly outperform human examiners.

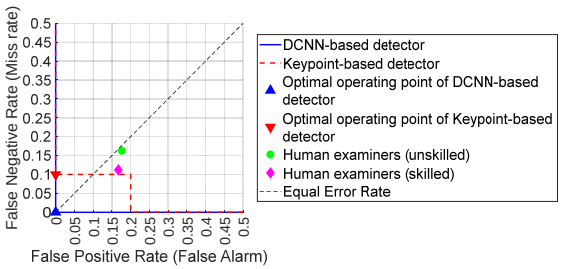


Figure 9: DET curves of Keypoint- and DCNN-based detectors along with error rates of humans.

6 CONCLUSION

Our study has once again confirmed the limited capability of human examiners to match unfamiliar faces as well as differentiate between morphed and genuine face photographs. We deliberately used automatically generated morphs with evident visual artifacts to demonstrate that human examiners can be deceived not only by professionally created manual morphs. Our experiment simulating a border control reveals the average MAR of 34.65% in the face matching scenario and the average FNR (miss rate) of 15.41% in the morphing detection scenario. In contrast, at least one of the algorithms used in our experiment is able to perfectly distinguish genuine and morphing trials in the matching experiment, or genuine and morphed images in the detection experiment, provided that a proper decision threshold has been selected. We understand that, due to the low number of samples, the error rates of algorithms resulting from our evaluation cannot be seen as reliable performance indicators and, therefore, cannot be generalized in any sense. We also understand that the error rates of our test participants might deviate from those of experienced border guards. Nonetheless, the experiment has shown clear trends and revealed general deficiencies of manual identity verification. Hence, we conclude that the manual processing of a document photograph constitutes the bottle neck of the concept of identity verification with a photo-ID, indicating the necessity for computer-aided support of photo-ID checking staff (e.g. border guards) in the field and at document issuing offices.

ACKNOWLEDGEMENTS

This work has been funded in part by the German Federal Ministry of Education and Research (BMBF) through the research programme ANANAS under the contract no. FKZ: 16KIS0509K. We thank Alexandra Koch, Dennis Siegel, Janine Zoellner, Kevin Michael Schott and Gina Marisa Seckendorf for the implementation of the experiment.

REFERENCES

Ferrara, M., Franco, A., Maltoni, D., 2016. On the Effects of Image Alterations on Face Recognition Accuracy. In *Bourlai, T. (ed.) Face Recognition Across the Electromagnetic Spectrum*, Springer, pp. 195-222.

FRONTEX, 2015. Best Practice Technical Guidelines for Automated Border Control (ABC) Systems.

Hancock, P., Bruce, V., Burton, A.M., 2000. Recognition of unfamiliar faces. *Trends in Cog. Sci. 4(9): 330-337*.

ICAO, 2004. Biometric Deployment of Machine Readable Travel Documents, TAG MRTDINTWG, May 2004.

ISO/IEC JTC1 SC17 WG3, 2018. Portrait Quality (Reference Facial Images for MRTD), Tech.Rep.2018.

King, D., Dlib C++ Library, Release notes, http://dlib.net/release_notes.html, *accessed 5.10.2018*

Kraetzer, C. et al., 2017. Modeling Attacks on Photo-ID Documents and Applying Media Forensics for the Detection of Facial Morphing, In *Proc. IH&MMSec '17*, pp. 21-32.

Luxand Inc., 2018. FaceSDK 6.5.1 Release Notes, https://www.luxand.com/facesdk/whatsnew/, *accessed 5.10.2018*

Makrushin, A., Neubert, T., Dittmann, J., 2017. Automatic generation and detection of visually faultless facial morphs, In *Proc. VISAPP'18*, pp. 39-50.

Makrushin, A., Wolf, A., 2018. An Overview of Recent Advances in Assessing and Mitigating the Face Morphing Attack. In *Proc.EUSIPCO'18*, pp1022-1026

Neubert, T., Makrushin, A., Hildebrandt, M., Kraetzer, C., Dittmann, J., 2018. Extended StirTrace Benchmarking of Biometric and Forensic Qualities of Morphed Face Images, *IET Biometrics 7(4):325-332*.

Phillips, P.J., O'Toole, A.J., 2014. Comparison of human and computer performance across face recognition experiments. *Image & Vis.Comp. 32:74-85*.

Raghavendra, R., Busch, C., 2017. Presentation Attack Detection Methods for Face Recognition Systems: A Comprehensive Survey, *ACM Comp. Surv. 50(1):8*.

Raghavendra, R. et al., 2017. Transferable Deep-CNN features for detecting digital and print-scanned morphed face images, *Proc.CVPRW'17*, pp.1822-1830

del Rio, J.S, Moctezuma, D., Conde, C., de Diego, I.M., Cabello, E., 2016. Automated border control e-gates and facial recognition systems, *Comp.& Sec.62:49-72*.

Robertson, D.J., Kramer, R.S.S., Burton, A.M., 2017. Fraudulent ID using face morphs: Experiments on human and automatic recognition, *PLoS ONE 12(3): e0173319*.

Robertson, D.J., et al., 2018. Detecting morphed passport photos: a training and individual differences approach. *Cognitive Research: Principles and Implications 3:27*.

Scherhag, U., Raghavendra, R., Raja, K.B., Gomez-Barrero, M., Rathgeb, C., Busch, C., 2017. On the Vulnerability of Face Recognition Systems: Towards Morphed Face Attacks, In *Proc IWBF'17*, pp. 1-6.

Seibold, C., Samek, W., Hilsmann, A., Eisert, P., 2017. Detection of Face Morphing Attacks by Deep Learning, In *Proc. IWDW'17*, pp. 107-120.

Seibold, C., Samek, W., Hilsmann, A., Eisert, P., 2018. Accurate and Robust Neural Networks for Security Related Applications Exampled by Face Morphing Attacks. *CoRR abs/1806.04265*.

White, D., Kemp, R.I., Jenkins, R., Matheson, M., Burton, A.M., 2014. Passport Officers' Errors in Face Matching. *PLoS ONE 9(8): e103510*.