Face Morphing Detection: Issues and Challenges

J. Merkle[1], C. Rathgeb[1,2], U. Scherhag[2], C. Busch[2], R. Breithaupt[3]
[1]*secunet Security Networks AG, Essen, Germany*
johannes.merkle@secunet.com, christian.rathgeb@secunet.com
[2]*Hochschule Darmstadt, Darmstadt, Germany*
ulrich.scherhag@h-da.de, christoph.busch@h-da.de
[3]*Federal Office for Information Security (BSI), Bonn, Germany,*
ralph.breithaupt@bsi.bund.de

Abstract: Recently, facial recognition systems have been found vulnerable to morphing attacks. In these attacks, the facial images of two (or more) individuals are combined (morphed) and the resulting morphed facial image is then presented during registration as a biometric reference. If the morphed image is accepted, it is likely that all individuals that contributed to the morphed facial image can be successfully authenticated against it. Morphing attacks thus pose a serious threat to facial recognition systems, in particular in border control scenarios, where the reference image is often provided in printed form by the applicant. This paper provides a rough overview of the current state-of-the-art methods for detecting morphed facial images, and discusses issues and challenges in the development and evaluation of morphing attack detection methods.

Keywords: face recognition; face morphing attacks; morphing attack detection; vulnerability analysis; issues and challenges

INTRODUCTION

Image morphing techniques can be used to combine information from two (or more) images into one image. Morphing techniques can also be used to create a morphed facial image from the biometric face images of two individuals, of which the biometric information is similar to that of both individuals. An example of a morphed facial image (hereinafter referred to as "morph") is shown in Figure 1.

Figure 1: Example of a morphed facial image. The morph was created with FantaMorph. On the left and right the contributing subjects are depicted and in the middle the resulting morph (image source: Hochschule Darmstadt, BSI).

In many countries, the facial image used for an electronic travel document is provided by the applicant either in analogue (i.e. print on paper) or digital form. Therefore, an *attacker* (e.g., a wanted criminal or a foreigner not eligible for entry to the Schengen area) could morph his face image with the face image of a similar looking *accomplice*, and the accomplice could apply for a passport or another electronic travel document with that image. It should be noted that morphed facial images look realistic and may be similar enough to both individuals to deceive human examiners [1][2]. This was showcased in Germany by members of the political activist group Peng! Kollektiv, who succeeded without any problem in applying for a passport with a morphed face image[1]. Both, the attacker and the accomplice can then be successfully verified against the morphed image so that the attacker can also use the electronic travel document issued to the accomplice to pass through an automatic border control (or even human inspections at border crossings). If more than two images are morphed, this usually reduces the attacker's chances of success if his characteristics are weaker in the resulting morph. The risk of the described *morphing attack* (MA) [3] is increased by the fact that realistic looking morphed facial images can be generated by unskilled persons. This can be done with the help of an easy-to-use morphing software for facial images, e.g., FantaMorph[2], which is either freely available or can be purchased at a reasonable price.

VULNERABILITY ANALYSIS

When analyzing the vulnerability of face recognition systems to MAs, it is obvious to augment the metrics for evaluation of presentation attacks, in which an attacker, for example, holds a photograph of another subject in front of the camera. The *Impostor Attack Presentation Match Rate* (IAPMR) [4] introduced in ISO/IEC 30107-3 represents a standardized metric for evaluating the impact of a presentation attack. The IAPMR is defined as follows: the proportion of impostor attack presentations species in which the target reference is matched in a full-system evaluation.

However, the disadvantage of the IAPMR metric for the evaluation of MAs is that it is calculated from individual attacks and therefore only reflects the probability of success of one of the subjects involved in the attack. In fact, however, two different scenarios can be relevant:

1. Only the attacker wants to be successfully authenticated by the face recognition system. In this scenario it is assumed that an accomplice was able to successfully apply for a passport, i.e. a human inspection of the morphed image was already overcome when the application was submitted. In such a scenario an asymmetric morphing of images, so that attacker and accomplice(s) contribute with different weights (a.k.a. alpha factors) to the morphed image, can be useful. An asymmetrical morphing can also be realized by procedures which morph the faces only in the inner area and the outer area (with forehead, hair, ears, neck) is taken only from one of the two initial images. It is usually assumed in the literature that the face of the accomplice contributes more to the morph than that of the attacker and that the outside area of the accomplice is used, because the risk of the picture being rejected during the application process is then lower.

1 Peng! Kollektiv, MaskID: https://pen.gg/de/campaign/maskid/
2 FantaMorph, Abrasoft: http://www.fantamorph.com/

However, since serious consequences (e.g., criminal prosecution[3]) are hardly to be expected in the case of a rejection in the application process, the reverse case, in which the accomplice is represented to a lesser extent in the morph, would also be conceivable. If only the attacker is to be successfully verified, the IAPMR can be used as a metric to evaluate the overall system's vulnerability. Care should be taken to ensure that the morphs used in the evaluation can at least overcome human inspections when presented by the accomplice, so that they are accepted when the application is made.

2. All individuals contributing to the morph want to be successfully authenticated against the morphed facial image. In such a scenario a symmetrical morphing of images is more realistic, i.e. attackers and accomplices contribute equally to the morphed image[4]. This scenario cannot be evaluated using the IAMPR and motivated the introduction of new evaluation metrics [5]. The comparison of a morphed facial image with a face image of a contributing subject is called a paired morph comparison. A MA is successful, if all involved subjects have been successfully verified. Hence, the minimum (for similarity values) or the maximum (for distance values) of all paired morph comparisons is of particular interest. Motivated by ISO/IEC 30107-3 [4], the *Mated-Morph-Presentation-Match-Rate* (MMPMR) is proposed in [5] to evaluate the effect of a MAs on the overall system.

MORPHING ATTACK DETECTION

In order to detect MAs, so-called *morphing attack detection* (MAD) techniques must be developed, which allow reliable differentiation between morphs and bona fide (i.e., genuine) facial images. If a potentially morphed facial image is detected in the course of an automatic border control, it can be inspected in a second step, e.g., by a border official, or the identity of the suspect can be checked using the fingerprints stored on the electronic passport. A particular challenge is the detection of analog morphs, i.e. after they have been printed and scanned, since many artefacts that indicate morphing can be lost due to the print-scan transformation. This is particularly relevant for passports from countries such as Germany, where an application with facial images in analogue form is still the rule.

DETECTION SCENARIOS

MAD procedures can be divided into two classes, see Figure 2, according to the scenario under consideration:

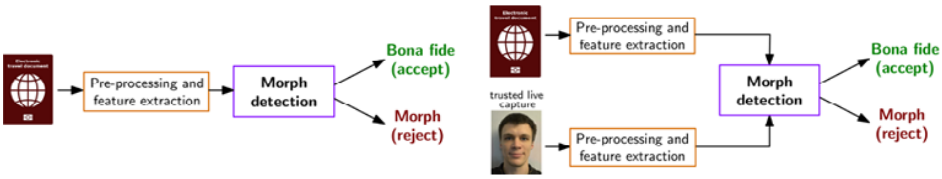

Figure 2: Morphing attack detection scenarios. Left: Single image Morphing Attack Detection, Right: Differential Morphing Attack Detection (image source: Hochschule Darmstadt).

3 It is also questionable whether an application for a passport with a morphed picture is a criminal offence, since even a morphed picture technically represents a photograph of the passport holder, which is clearly what is required, for instance by the German Passport Act.

4 However, the outside area can be taken over by only one subject to avoid possible morphing artefacts in that area.

- *Single image MAD*: These approaches examine a single face image, for example, when checking the authenticity of a passport without reference or directly when applying for a passport, and check whether it has been morphed. For this purpose, the image is examined for potential traces of a morphing process. This class of MAD procedures is also known as no-reference MAD or forensic MAD.

- *Differential MAD*: These procedures compare the potentially morphed reference image with a trusted probe image, e.g., a live image from an eGate for automatic border control. This class of MAD procedures is also referred to as image pair-based MAD.

Basically, the two approaches differ in that single image MAD approaches aim to detect certain artefacts induced by the morphing process (e.g., "ghost artefacts" in which structures of the original images overlap), while differential MAD methods analyze the features of the potentially morphed facial image and the live image of a face, e.g. by estimating difference vector between both feature vectors. It can be assumed that carefully created morphs contain only a few recognizable artefacts (if any), which after a print-scan process (i.e., when providing an analog facial image) are probably very difficult to detect. Single image MAD procedures can depend heavily on the training data used and can only detect the artefacts learned during training. This can greatly limit the generalizability of these methods. For these reasons, differential MAD procedures are generally to be seen as more promising.

In recent years, numerous approaches for the automated detection of MAs have been presented. A detailed overview is given in [3]. The majority of works is based on the single image scenario. Despite promising results reported in many studies, the reliable detection of morphed facial images is still an open research task. In particular, the generalizability and robustness of the published approaches could not yet be proven. The results are hardly comparable and comprehensible. The vast majority of publications use internal databases of the respective research groups for training and testing. In addition, different evaluation metrics are used in the publications, and some even state error rates of zero without specifying the number of samples. Since most implemented MAD procedures are not made publicly accessible, no comparative independent evaluation of the detection performance is possible (without cooperation with the respective authors).

Furthermore, most publications only use images from a single database and morphs generated with a single algorithm for training and testing, so that the generalization capability of the methods cannot be assessed across different databases and morphing methods. In publications on differential MAD, the comparison images used often show a low variance with respect to poses, facial expressions and illumination and are usually produced shortly after the reference image - in real scenarios such as border control, a much higher variance is to be expected. In addition, most studies neglect the probable application of image post-processing techniques by an attacker, such as subsequent image sharpening, and the print-scan transformation.

SINGLE IMAGE MAD APPROACHES

The single-image MAD approaches can be categorized into three classes: Texture descriptors, e.g., in [6], forensic image analysis, e.g., in [7], and methods based on deep neural networks, e.g., in [8]. These differ in the artefacts they can potentially detect. A brief overview is given in Table 1.

Table 1: Categories of singe image MAD approaches.

Category	Analyzed artefacts
Texture descriptors	Smoothened skin texture, ghost artefacts/ half-shade effects (e.g., on pupils, nostrils), distorted edges, offset image areas
Forensic image analysis	Sensor pattern noise, compression artefacts, inconsistent illumination or color values
Deep-learning approaches	All possible artefacts learned from a training dataset

DIFFERENTIAL MAD APPROACHES

Differential MAD can be categorized into approaches that perform a biometric comparison directly with the two facial images, e.g., in [9], and algorithms that attempt to reverse the (potential) morphing process, e.g., in [10]. In the former category, features from both face images, the potentially morphed facial image and the probe image, are extracted and then compared. The comparison of the two feature vectors and the classification as bona fide comparison or MA is usually done using machine learning techniques. By specifically training these procedures for the recognition of MAs, they can - in contrast to facial recognition algorithms - learn to recognize specific patterns within the differences between the two feature vectors for these attacks. This has already been demonstrated for features derived from general purpose texture descriptors. While training a deep neural network from scratch in order to learn discriminative features for MAD requires a high amount of training data, pre-trained deep networks can be employed.

The second type of differential MAD procedure aims at reversing the morphing process in the reference image ("de-morphing") by using a probe image. If the reference image was morphed from two images and the probe image shows a person contributing to the morph (the attacker), the face of the accomplice would ideally be reconstructed, which would be rejected in a subsequent comparison with the probe image using biometric face recognition; if, on the other hand, a bona fide reference image is available, the same subject should still be recognizable after the reversal of a presumed morph process with the probe image, and thus the subsequent comparison of the facial recognition process should be successful.

MAD BASED ON DEEP FACE REPRESENTATIONS

For both single image MAD and differential MAD, a straightforward approach is to train a classifier on deep features computed by existing convolutional neural networks (CNNs) for

biometric face recognition. The advantage of this approach is that it benefits from the strength of CNNs to extract relevant features from image data but does not require the large amount of data typically necessary to train a CNN. While the features extracted by face recognition networks have not been trained to detect morph attacks, at least in the differential scenario, they might still be very useful for MAD: As the morphed face image does not only contain biometric features of the attacker but also those of the accomplice, its deep face features should, at least in certain aspects, considerably deviate from those detected in the probe image. The vulnerability of face recognition networks to morph attacks does not necessarily imply that the features extracted by those are not eligible for MAD but can also be explained by an inaptly chosen classification method (which is typically based on simple geometric distances). Thus, one can hope that a new classifier trained for MAD on deep face features may be able to recognize the characteristic differences in the features between morphs and probe images.

In [11], deep face representations, i.e., VGG-Face16 and VGG-Face2, have been employed to train machine learning-based classifiers for single-image MAD. Promising detection rates have been reported in the presence of printing/scanning and heterogeneous image sources.

In a preliminary study of the authors, conducted in the course of the FACETRUST project, deep face features of both commercial and open source face recognition systems were employed to develop differential MAD. Deep face representations extracted from reference and probe images were combined, e.g., by element-wise subtraction or concatenation, and the resulting vectors were then used to trained machine learning-based classifiers for differential MAD.

The following conclusions regarding performance/generalizability are reached:

- *Detection performance*: the detection performances achieved are promising and highly robust with respect to image post-processing, i.e., image compression, image resizing and even print-scan transformation. This is a clear advantage over MAD based on texture descriptors, which is typically quite sensitive to post-processing, particularly in more challenging scenarios. Moreover, in some cases it turned out to be favorable to perform training on digital images, which have not been printed and scanned, to obtain improved detection rates even for scanned images.

- *Heterogeneous morphing algorithms*: morphs generated by morphing algorithms which produce obvious artefacts, e.g., clearly visible ghost artefacts, were generally detected with higher accuracy. Furthermore, the recognition performance slightly degrades if training and evaluation sets contain morphs generated by different morphing algorithms.

- *Heterogeneous databases*: if training and testing is conducted on heterogeneous face image databases which contain face images with different conditions, e.g., variations in pose and lightning, detection performance is negatively affected. On databases obtained from subsets of the publicly available FERET and the FRGCv2 face database, experiments revealed higher detection accuracy on the FERET subset in which probe images only contain slight variations in expression and pose as opposed to the FRGCv2 subset, which additionally comprises probe images with variations in lightning and focus. It can be concluded that

strong variations in lightning and focus of probe images represent especially challenging conditions for differential MAD.

- *Machine learning-based classifiers*: among the tested machine learning-based classifiers, i.e., AdaBoost, Gradient Boosting, Random Forest and Support Vector Machine (SVM), SVM-based classifiers generally revealed most competitive detection performance across the vast majority of conducted experiments.

- *Commercial vs. open-source*: while commercial face recognition algorithms frequently outperform corresponding open-source implementations, this is not necessarily the case for MAD. Precisely, for the task of MAD, deep face representations obtained from open-source algorithms, e.g. FaceNet or ArcFace, might be better suited, compared to deep features extracted by commercial face recognition systems.

ISSUES AND CHALLENGES

In research on MAD, there are various open questions and challenges:

Evaluation metrics: Even though initial efforts have already been made to introduce them, standardized metrics for evaluating the performance of MAD procedures are not yet available; these should be defined uniformly (ideally as an international standard) and applied in publications on MAD procedures in order to enable a meaningful comparison of the proposed approaches.

Evaluation protocols: To obtain reproducible and statistical significant results performance evaluations of proposed MAD approaches should be transparent and based on sufficient data. Used face databases must be split into subject-disjoint sets for training and evaluation. Reporting the used number of sample and conducted amount of comparisons is essential in order to interpret obtained results in a meaningful way.

Generalizability of MAD approaches: The majority of the MAD methods published so far - in particular the single image MAD methods - aim at the detection of artefacts that can easily be avoided, e.g., clearly visible ghost artefacts, double compression artefacts and changed image noise patterns. Hence, reported detection rates tend to be over-optimistic. In contrast, research should focus on the development of MAD methods that detect artefacts that are difficult to avoid. In addition, MAD approaches are, like any classification task, susceptible to overfitting to training data. Therefore, when evaluating MAD approaches, images of which source and properties differ from those of the training data, i.e., images from other databases and morphs created with other techniques, should be employed.

For border control scenarios, MAD techniques need to be robust against print-scan transformations, resizing and strong compression of reference images. Similarly, in the case of differential MAD, considerable variance of illumination, background, pose, appearance (hair, beard, glasses, etc.) and aging (up to 10 years for passports) can be expected in probe images. In order to be applicable to these scenarios, MAD approaches should be trained and evaluated on images exhibiting these characteristics.

Unfortunately, post-processing steps applied to reference images like printing/scanning and strong image compression have been found to cause drastic drops in the detection performance at least for single image MAD, since artefacts caused by morphing vanish in the post-processed reference. In order to reduce this issue in the long term, responsible authorities should raise the requirements for image quality, resolution and size of face images to be stored in electronic travel documents. Eventually, the susceptibility of the passport issuance processes can be eliminated by using live enrolment stations.

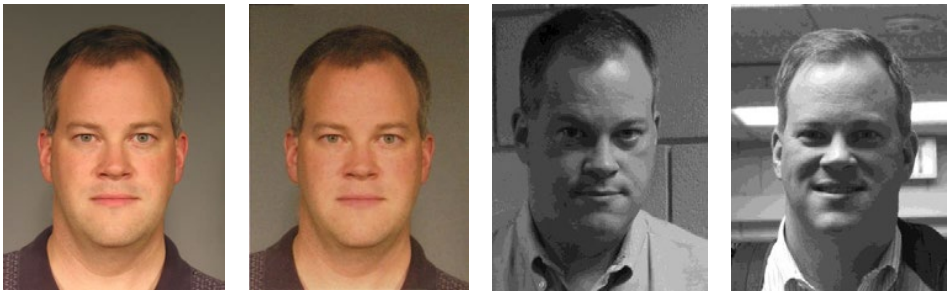

Figure 3: From left to right: original reference; reference printed, scanned (300 dpi), resized (360x465 pixels) and compressed (JPEG 2000, 15KB); probe with slight rotation; probe with changing expression and variation in illumination.

Databases: Currently, the publicly available facial image databases do not represent the characteristics and variance of real-world scenarios. To the authors' knowledge, there is no public database containing a large number of printed and scanned facial images. Furthermore, there is no database comprising face images which fulfill the conditions of reference and probe images needed to simulate a realistic boarder control scenario, i.e., containing both images conforming to the ICAO specifications for passport photographs and images resembling all variations (in particular aging) to be expected for live images in a border control. Figure 3 depicts face images taken from the FRGCv2 database which reflect at least some of the variance expected in a real border control scenario. In addition, there is just one database with morph images of good quality that has been made available[5], and the creation of morphs of high quality is still laborious with publicly available tools.

In order to overcome this issue, border control agencies could collect large databases with images that resemble the characteristics of images typically met in border control scenarios. These images should comprise bona fide reference images taken in accordance with ICAO requirements [12] as well as high-quality morphs of these (created with various methods). To all reference images realistic post-processing steps (e.g., printing and scanning, resizing to approx. 400x500 pixels and JPEG-2000 compression to 15KB) should be applied. The database should also contain corresponding probe images with realistic distribution of illumination, pose, appearance and aging. It should also be taken into account that in morph attacks, the variance between reference and probe is likely to be smaller than for bona fide authentication attempts. Ideally, such database would be made available to researchers for the development

5 https://www.linkedin.com/pulse/new-face-morphing-dataset-vulnerability-research-ted-dunstone

and evaluation of MAD methods. If operational data cannot be made available due to data protection legislation, images could be captured with volunteers under realistic conditions, e.g., using automatic border control gates.

The detection performance of differential MAD approaches can be influenced by the quality of the captured probe image. It is well-known that high recognition performance can only be achieved if the quality of the captured facial data is sufficient. As stressed in a recent study [13] by the Joint Research Centre (JRC) of the European Union, algorithms must be incorporated to ensure a robust determination of the face image quality.

Transparency: In scientific publications, the MAD procedures are usually presented in a way that they cannot easily be re-implemented by third parties without considerable effort while resulting re-implementations hardly achieve comparable recognition performance. Implementations of MAD procedures should therefore be made publicly available in order to guarantee the reproducibility of results that were achieved on public data. It is expected that the planned benchmark program of the National Institute of Standards and Technology (NIST) [14] will enable a quantitative comparison of published approaches in the near future. Border control agencies could support this program by providing realistic image data or information on the characteristics and variance of the images to be expected in border control scenarios.

SUMMARY

Morph attacks pose a high security risk to modern facial recognition systems in particular for border control. To counteract this, reliable methods for morph attack detection must be developed. Various research groups from the fields of image processing and biometrics have recently published scientific papers on this topic, and several publicly funded research projects are currently dealing with this problem. However, research in this field is still in its infancy and does typically not address the variance of the image data available in border control scenarios. The development of MAD approaches that are effective and robust in real-world scenarios will require a considerable amount of future research as well as close collaborations with border guard agencies.

ACKNOLEDGEMENTS

This work was partially supported by FACETRUST project of the Federal Office for Information Security (BSI).

REFERENCES

[1] M. Ferrara, A. Franco and D. Maltoni, „On the Effects of Image Alterations on Face Recognition Accuracy“, in Face Recognition Across the Imaging Spectrum, Springer International Publishing, 2016.

[2] J. D. Robertson, A. G. Mungall, D. Watson, A. K. Wade, J. S. Nightingale and S. Butler, „Detecting morphed passport photos: a training and individual differences approach“, Cognitive Research: Principles and Implications, 2018.

[3] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch. Face recognition systems under morphing attacks: A survey. IEEE Access, 7:23012–23026, 2019.

[4] ISO/IEC JTC1 SC37 Biometrics, ISO/IEC IS 30107-3:2017, IT – Biometric presentation attack detection – Part 3: Testing and Reporting, 2017.
[5] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. J. Veldhuis, L. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, R. Breithaupt and R. Ramachandra, „Biometric Systems under Morphing Attacks: Assessment of Morphing Techniques and Vulnerability Reporting", in Proceedings of the 2017 International Conference of the Biometrics Special Interest Group (BIOSIG), 2017.
[6] R. Ramachandra, K. B. Raja and C. Busch, „Detecting morphed face images", in Proceedings of the 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), 2016.
[7] C. Kraetzer, A. Makrushin, T. Neubert, M. Hildebrandt and J. Dittmann, „Modeling Attacks on Photo-ID Documents and Applying Media Forensics for the Detection of Facial Morphing", in Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security - IHMMSec '17, 2017.
[8] C. Seibold, W. Samek, A. Hilsmann and P. Eisert, „Detection of Face Morphing Attacks by Deep Learning", in Digital Forensics and Watermarking, 2017.
[9] U. Scherhag, C. Rathgeb and C. Busch, „Towards detection of morphed face images in electronic travel documents", in Proceedings of the 13th IAPR Workshop on Document Analysis Systems (DAS), 2018.
[10] M. Ferrara, A. Franco und D. Maltoni, „Face Demorphing", IEEE Transactions on Information Forensics and Security, 2018.
[11] M. Ferrara, A. Franco, D. Maltoni: „Face morphing detection in the presence of printing/scanning and heterogeneous image sources", arXiv:1901.08811, 2019.
[12] International Civil Aviation Organization, ICAO Doc 9303, Ma-chine Readable Travel Documents - Part 9: Deployment of Biometric Identification and Electronic Storage of Data in MRTDs (7th edition), 2015.
[13] J. Galbally, P. Ferrarra, R. Haraksim, A. Psyllos and L. Beslay, "Study on Face Identification Technology for its Implementation in the Schengen Information System", Publications Office of the European Union, 2019.
[14] M. Ngan, P. Grother and K. Hanaoka, „Performance of Auto-mated Facial Morph Detection and Morph Resistant Face Recognition Algorithms", National Institute of Standards and Technology (NIST), 2018.