VULNERABILITY ANALYSIS OF FACE MORPHING ATTACKS FROM LANDMARKS AND GENERATIVE ADVERSARIAL NETWORKS

Eklavya Sarkar, Pavel Korshunov, Laurent Colbois, Sébastien Marcel

Idiap Research Institute, Martigny, Switzerland
{eklavya.sarkar, pavel.korshunov, laurent.colbois, sebastien.marcel}@idiap.ch

ABSTRACT

Morphing attacks is a threat to biometric systems where the biometric reference in an identity document can be altered. This form of attack presents an important issue in applications relying on identity documents such as border security or access control. Research in face morphing attack detection is developing rapidly, however very few datasets with several forms of attacks are publicly available. This paper bridges this gap by providing a new dataset with four different types of morphing attacks, based on OpenCV, FaceMorpher, WebMorph and a generative adversarial network (StyleGAN), generated with original face images from three public face datasets. We also conduct extensive experiments to assess the vulnerability of the state-of-the-art face recognition systems, notably FaceNet, VGG-Face, and ArcFace. The experiments demonstrate that VGG-Face, while being less accurate face recognition system compared to FaceNet, is also less vulnerable to morphing attacks. Also, we observed that naïve morphs generated with a StyleGAN do not pose a significant threat.

Index Terms— Biometrics, Face Recognition, Vulnerability Analysis, Morphing Attack, StyleGAN 2

1. INTRODUCTION

After Ferrara *et al.* [1] showed that by using a morphed photo of two different people an adversary can circumvent passport registration process, morphing attacks and how to detect them received a lot of attention from academic, industrial, and security communities. The vulnerability of state-of-the-art (SOTA) face recognition systems (FR) and the threat such vulnerability poses to the security systems relying on recognition technologies led to the explosion of research work in this area.

Most of the work related to morphing attacks (MAs) focuses on their detection. Recently proposed techniques for morphing attack detection (MAD) include methods based on *so called* classical approaches using local binary patterns (LBP) and support vector machines (SVM) [2], approaches rooted in image forensics that rely on photo response non uniformity (PRNU) function [3], deep neural networks specifically trained to detect morph images [4], and FR systems themselves serving as feature extractors for an support vector machine (SVM) classifier [5]. The National Institute of Standards and Technology (NIST) is now conducting independent evaluations of MAD technologies [6].

Because the researchers are mainly focused on detecting MAs, some related questions suffer from lack of attention: i) very few public databases of morphed images are available and ii) little is reported on whether the latest SOTA face recognition systems remain vulnerable to the typical morphing attacks.

It is common for MAD systems to be evaluated on morphed images that are privately generated using real data from public datasets [5, 7, 8] and open source tools for face morphing such as OpenCV [9]. Very few databases are publicly available with a few exceptions, notably, the Advanced Multimedia Security Lab's (AMSL) Face Morph Image dataset [10], which is the exception that proves the common practice. Also, the advent of generative adversarial networks (GANs) opens up new possibilities for generating possibly more realistic *deep morphing* attacks [11], which are not yet well explored.

Therefore, in this paper, we aim to fill these gaps by proposing several publicly available datasets with morphed images, including with the latest GAN-based morphs, and an extensive vulnerability assessment of the SOTA face recognition systems. Using open source morphing implementations OpenCV-based [9] and FaceMorpher [12], online tool WebMorph [13], and the latest StyleGAN 2 [14], we generated morphed images using real faces from publicly available FERET [15] and FRGCv2 [16] datasets and also extend AMSL Face Morph Image [10] dataset.

We use these three different datasets and different types of morphs to assess the vulnerability of SOTA systems such as FaceNet [17] and VGG-Face [18], as well as ArcFace [19], which is used in some of the latest morphing attacks vulnerability studies [8], and baseline "classical" systems based on Gabor filters [20] and Inter-Session Variability (ISV) modelling [21]. Therefore, besides providing several morphed images to public, this paper is also an important milestone in our understanding of where the current state of the art morph generation algorithms and FR systems are at.

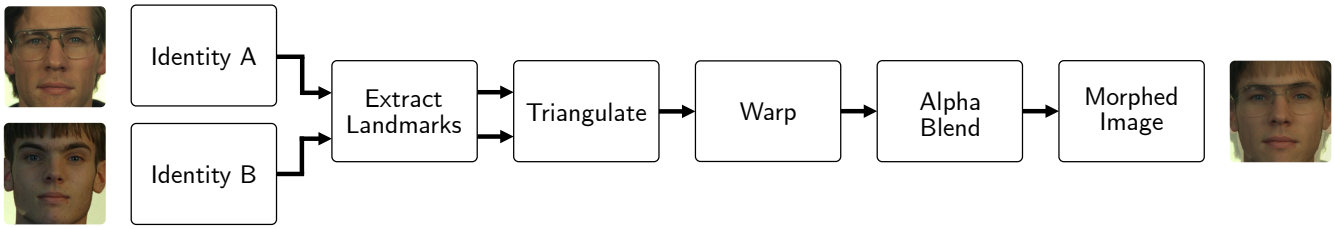

Fig. 1. Morphed image generation pipeline for landmark-based methods.

To allow researchers to verify, reproduce, and extend our work, we provide all scripts for generation of morphed images for FERET, FRGC, and FRLL (the dataset ASML Face Morph Image set is based on) datasets, FR systems used, and the scripts for their vulnerability assessment as open source[1]. Please note that due to the licensing restrictions of the original FERET, FRGC, and FRLL datasets, we are unable to provide the morphed images directly, but we do provide easily installable and well documented code that can be used to regenerate all the morphs we created, given one has legally obtained the copies of the original image datasets.

2. MORPH GENERATION

In this section, we present the datasets with bona fide faces and the different tools, including the one based on GANs, used to generate the morphing images for the vulnerability experiments.

2.1. Datasets

We used FERET [15], FRGC v2.0 [16], and Face Research Lab London (FRLL) [22] datasets of facial images to generate the morphs. FERET and FRGC were selected because they are *de facto* the standard datasets commonly used in papers on morphing attack detection [8, 23] and they have large number of images of different identities. We also used FRLL dataset, because the only practically available dataset of morphed images, AMSL Face Morph Image dataset generated by Neubert *et al.* [10] used the original faces from FRLL dataset. Also, FRLL dataset is a great choice to use for creating morphing attacks, because it contains close-up frontal face images of very high visual quality and 1350×1350 resolution, shot under *uniform* illumination with large varieties in ethnicity, pose, and expression. Each face is annotated using 189 facial landmarks, which is notably a very high number, as typical landmarks detectors provide no more than 68-70 landmarks. The main limitation of FRLL dataset, compared to FERET and FRGC datasets, is the limited number 102 of different identities with 53 males and 49 females.

For each dataset, we select bona fide (or original) face pairs for morph generation by following the existing protocols used in previous work. For FERET and FRGC, we follow the protocols used in the work by Scherhag *et al.* [8] that were kindly provided by the authors (though they were not able to provide any morphed images). For FRLL dataset, we follow the protocols used in AMSL Face Morph Image dataset by Neubert *et al.* [10]. Using these protocols (essentially, which facial image pairs to morph), we generated morphs with the following morphing tools: OpenCV, FaceMorpher, WebMorph, and StyleGAN 2, which are described in Section 2.2.

2.2. Morphing Tools

As morphing tools, we selected two commonly used open source face morphing algorithms, OpenCV-based [9] and FaceMorpher [12], web-based open source morphing tool called WebMorph [13], and the algorithm adapted from the recently proposed StyleGAN 2 implementation [14].

2.2.1. Landmarks-Based Morphs

The **OpenCV** algorithm is an adaptation of an open-source implementation [9] used for morphing faces using 68-point annotator from Dlib library [24]. Face landmarks are obtained for each of the two bona fide source images and are used to form Delaunay triangles, which are in-turn warped and alpha blended.

FaceMorpher is also an open-source [12] similar to OpenCV landmark-based morphing algorithm, but with the STASM [25] landmark detector instead. Both algorithms create morphs with noticeable ghosting artefacts for all three datasets, as the region outside the area covered by these landmarks is simply averaged.

WebMorph [13] is an online landmark-based morphing tool created by the FRLL dataset providers, which requires 189 landmarks that are available only in FLLR dataset, to generate morphed images with better alignment and of an overall higher visual quality. Ghosting artefacts are still visible and prominent around the hair and neck area, but are noticeably improved around the ears. As this tool works exclusively with the annotation files of FRLL dataset, we were not able to generate the same types of morphs for FERET and FRGC datasets.

[1] available upon publication

(a) Identity A (b) OpenCV (c) FaceMorpher (d) StyleGAN2 (e) WebMorph (f) CombinedMorph (g) Identity B

Fig. 2. Different types of generated morphed images on the FRLL dataset after initial preprocessing.

The morphed images in the AMSL dataset, which were generated by Neubert *et al.* [10] from bona fide images of FRLL dataset using the private Combined Morphs tool, contain very realistic morphs with virtually no ghosting artefacts, even around the hair and neck area, because of the additional poisson image editing. Unlike the other tools, this proprietary technique generates *two* unique morphed images for every pair of source bona fide images. Since Combined Morphs is a proprietary tool, we were unable use it for the generation of the same morphs for the FERET and FRGC datasets.

2.2.2. Generative Adversarial Network-Based Morphs

Following the advances in generative adversarial networks (GANs), there were attempts to generate morphed images using a GAN instead of landmark-based methods [11, 26]. In this paper, we adapted the latest **StyleGAN 2** [14] to develop a morphing algorithm which can generate high resolution realistic looking faces with no noticeable artifacts. The StyleGAN 2 was pre-trained on the FFHQ dataset introduced in [27].

The faces are cropped to obtain the same landmark alignment as in the FFHQ dataset. The images are then projected into the W space of StyleGAN 2 by optimizing the input latent style vector that is fed to the generator network, such that it minimizes the perceptual loss between the generated and real image [14]. Once an associated latent vector has been computed for each of the source images, morphs can be generated by linearly interpolating between two latent vectors, and feeding the interpolated vector back into the generator.

This technique yields very realistic looking morphs without visual artefacts, however, since StyleGAN does not have any information about the identities in bona fide images, there is no guarantee that the resulted morph is actually a blen of these identities (see the example in Figure 2(d) for an idea).

StyleGAN 2 also requires the projected images to be at a high resolution (1024x1024 after cropping), and works better with an uniform background, which makes the FRLL dataset particularly appropriate. A side not observation of using StyleGAN 2 for generating morphs is that it is equally easy to generate high-quality morphs for smiling expressions as it is for the neutral faces, which is not possible with typical landmark-based tools.

2.3. Generation

Using morphing tools presented in Section 2.2, we generate three sets of morphs each consisting of $1'222$, 529, and 964 images for FRLL, FERET, and FRGC datasets respectively. For FRLL, we also generate one additional set of morphs using WebMorph tool. Please note that the morphs for FERET and FRGC datasets are generated using the same protocols used in [8], while the morph generation protocol defined in ASML Face Morph Image dataset was used in case of FRLL.

3. EVALUATION PROTOCOL

3.1. Face Recognition Systems

To evaluate vulnerability of face recognition against morphing attacks, we used publicly available pre-trained FaceNet, ArcFace, and VGG-Face architectures. We used the last fully connected layers of these networks as features and the cosine distance as a classifier. For a given test face, the confidence score of whether it belongs to a reference model is the cosine distance between the average reference feature vector and the feature vector of a test face. These systems are the state of the art recognition systems with Facenet showing 99.63% [17], ArcFace – 99.53 [19], and VGG-Face – 98.95% [18] accuracies on the labeled faces in the wild (LFW) dataset.

We also used two "classical" baselines: i) Gabor jet implementation from LBP features [20] and ii) ISV-based face recognition [21]. DCT features computed on overlapping blocks of 40x40 were used for the ISV-based system of 512 Gaussian mixture models (GMMs) and 160 dimensional subspace, which was pre-trained on the MOBIO [28] dataset.

3.2. Evaluation Metrics

In a the verification process, a user attempting to authenticate presents a biometric probe and a claimed identity, and can be classified into one of the following three categories. *A) Genuine user* (BF): probe and claimed identity both correctly belong to the user. *B) Zero-effort impostor* (BF): probe belongs to the user, but the claimed identity corresponds to a different enrolled user. *C) Morph attack impostor* (MA): probe matches the claimed identity but does not correspond to the user.

The *verification* performance is typically evaluated with the following metrics. *A) False Match Rate (FMR)* [23]: proportion of zero-effort impostors that are falsely authenticated. *B) False Non-Match Rate (FNMR)* [23]: proportion of genuine users which are falsely rejected. *C) Mated Morph Presentation Match Rate (MMPMR)* [29]: proportion of morphs attacks impostors accepted by the face recognition system.

3.3. Evaluation scenarios

For the FERET and FRGC datasets, we adopted the same evaluation scenarios used in [8]. For the FRLL dataset, we defined our own evaluation protocols due to the lack of publicly available protocols for FRLL.

In general, there are two main scenarios under which a face recognition system is evaluated: a bona fide (BF) scenario where both the reference and probes images as genuine, so there are no attacks and the system is assumed to perform under the conditions it was designed for; and the morphing attack (MA) scenario when morphs are introduced to the face recognition with a malicious intent to spoof the recognition. There are also two variants of MA scenario, when a morphed image can be either used as a reference, i.e., FR system is hijacked during enrollment process (typical morphing attack scenario), or a morphed image is used as a probe, which is similar to presentation attack scenario. The number of reference and probe images for each evaluation scenario is summarized in Table 1.

Also note that we did not split datasets into training, development, and test subsets but used each whole dataset one test set. The reason for this is that all recognition systems we used were pre-trained on other databases, so there is no need for the training set, and we choose the decision threshold to compute MMPMR value for MA scenario based on FMR value computed in the bona fide scenario, so there is no need for a development set.

Table 1. Number of images in different evaluation scenarios.

Dataset	Morphs as	BF	MA	Impostors
FRLL	References	91	584	1,984
	Probes	584	91	4,153
FERET	References	529	791	418,439
	Probes	791	529	418,439
FRGC	References	3,298	964	1,698,384
	Probes	964	3,298	1,698,384

4. EXPERIMENTAL RESULTS

Table 2 summarizes the results of vulnerability assessment of the face recognition systems described in Section 3.1 under different morphing attack scenarios (see Section 3.3 for details). The MMPMR metric (see Section 3.2 for details) is calculated by setting the decision threshold at FMR=0.1% in the bona fide scenario.

Table 2. MMPMR @ FMR = 0.1% (morphs as references — morphs as probes) [%]

Dataset	FRS	OpenCV	FaceMorpher	StyleGAN2	WebMorph	AMSL
FRLL	FaceNet	83.3 — 72.0	64.5 — 68.2	5.9 — 11.0	82.7 — 70.8	89.2 — 92.5
	ArcFace	59.8 — 73.8	57.6 — 75.3	9.8 — 18.3	60.9 — 73.8	58.0 — 79.4
	VGG	39.7 — 48.6	23.4 — 47.1	3.0 — 9.1	38.2 — 52.2	65.7 — 89.8
	Gabor	87.2 — 100.0	83.9 — 99.4	11.8 — 37.9	85.4 — 100.0	86.3 — 99.9
	ISV	59.8 — 97.8	56.1 — 96.1	9.2 — 43.6	59.5 — 97.4	55.3 — 99.9
FERET	FaceNet	41.1 — 40.6	39.9 — 40.3	1.6 — 1.3	N/A	N/A
	Arcface	34.6 — 35.2	34.1 — 34.8	2.4 — 2.5	N/A	N/A
	VGG	22.0 — 21.0	20.5 — 18.3	2.0 — 1.5	N/A	N/A
	Gabor	66.6 — 90.9	63.7 — 88.5	1.3 — 40.8	N/A	N/A
	ISV	44.8 — 58.4	42.6 — 56.5	2.7 — 3.4	N/A	N/A
FRGC	FaceNet	6.9 — 5.9	7.0 — 5.7	1.0 — 0.7	N/A	N/A
	Arcface	11.9 — 10.8	12.1 — 11.2	0.5 — 0.4	N/A	N/A
	VGG	5.5 — 4.5	5.1 — 4.8	0.7 — 0.4	N/A	N/A
	Gabor	7.1 — 80.8	6.7 — 81.0	0.6 — 75.8	N/A	N/A
	ISV	4.2 — 6.5	3.5 — 6.2	0.6 — 0.6	N/A	N/A

The results in Table 2 illustrate two important observations:

1. The more accurate face recognition systems are the more vulnerable they are to the morphing attacks regardless of whether they are used as references or as probes, which is also in line with the observations made for presentation attacks [30]. This trend is especially evident when we compare a more accurate and deeper FaceNet architecture with VGG-Face for all databases and types of morphs.

2. The morphs generated with StyleGAN 2, one of the latest generative adversarial network, do not pose significant threats to the state of the art recognition systems. A slight elevation in the MMPMR error for StyleGAN2 morphs of FRLL dataset indicates that high quality original images may lead to slightly more accurate morphs.

5. CONCLUSION

The paper presents an extensive vulnerability assessment of the state of the art face recognition systems based on VGG-Face, ArcFace, and FaceNet neural network models on three image databases with five different morphing attacks, one of which was created using StyleGAN 2. The experiments demonstrate that a more accurate face recognition system FaceNet is more vulnerable to the morphing attacks and that GAN-based morph do not yet pose a significant threat to modern recognition systems. However, if one would introduce an identity loss into StyleGAN-based morph generation to ensure that the identities of both bona fide inputs are preserved in the resulted morph, then the GAN-based morphs may become highly threatening to face recognition and this would be an interesting future direction.

6. REFERENCES

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

[2] L. Spreeuwers, M. Schils, and R. Veldhuis, "Towards robust evaluation of face morphing detection," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1027–1031.

[3] U. Scherhag, L. Debiasi, C. Rathgeb, C. Busch, and A. Uhl, "Detection of face morphing attacks based on prnu analysis," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 1, no. 4, pp. 302–317, 2019.

[4] Clemens Seibold, Wojciech Samek, Anna Hilsmann, and Peter Eisert, "Accurate and robust neural networks for face morphing attack detection," *Journal of Information Security and Applications*, vol. 53, pp. 102526, 2020.

[5] L. Wandzik, G. Kaeding, and R. V. Garcia, "Morphing detection using a general- purpose face recognition system," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1012–1016.

[6] Mei L. Ngan, Patrick J. Grother, Kayee K. Hanaoka, and Jason M. Kuo, "Face recognition vendor test (frvt) part 4: Morph - performance of automated face morph detection," Tech. Rep., National Institute of Standards and Technology, 2020.

[7] N. Damer, A. M. Saladié, A. Braun, and A. Kuijper, "Morgan: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2018, pp. 1–10.

[8] U. Scherhag, C. Rathgeb, J. Merkle, and C. Busch, "Deep face representations for differential morphing attack detection," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3625–3639, 2020.

[9] Satya Mallick, "Face morph using opencv — c++ / python," March 2016.

[10] T. Neubert, A. Makrushin, M. Hildebrandt, C. Kraetzer, and J. Dittmann, "Extended stirtrace benchmarking of biometric and forensic qualities of morphed face images," *IET Biometrics*, vol. 7, no. 4, pp. 325–332, 2018.

[11] Pavel Korshunov and Sébastien Marcel, "Vulnerability of face recognition to deep morphing," in *International Conference on Biometrics for Borders*, oct 2019, pp. 1–5.

[12] Alyssa Quek, "Facemorpher," January 2019.

[13] Lisa DeBruine, "debruine/webmorph: Beta release 2," Jan. 2018.

[14] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8107–8116.

[15] P.Jonathon Phillips, Harry Wechsler, Jeffery Huang, and Patrick J. Rauss, "The feret database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing*, vol. 16, no. 5, pp. 295 – 306, 1998.

[16] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, Jin Chang, K. Hoffman, J. Marques, Jaesik Min, and W. Worek, "Overview of the face recognition grand challenge," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 1, pp. 947–954 vol. 1.

[17] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823.

[18] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.

[19] Jiankang Deng, Jia Guo, Xue Niannan, and Stefanos Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *CVPR*, 2019.

[20] Manuel Günther, Dennis Haufe, and Rolf P. Würtz, "Face recognition with disparity corrected gabor phase differences," in *Artificial Neural Networks and Machine Learning – ICANN 2012*, Alessandro E. P. Villa, Włodzisław Duch, Péter Érdi, Francesco Masulli, and Günther Palm, Eds., Berlin, Heidelberg, 2012, pp. 411–418, Springer Berlin Heidelberg.

[21] R. Wallace, M. McLaren, C. McCool, and S. Marcel, "Inter-session variability modelling and joint factor analysis for face authentication," in *2011 International Joint Conference on Biometrics (IJCB)*, 2011, pp. 1–8.

[22] Lisa DeBruine and Benedict Jones, "Face Research Lab London Set," 5 2017.

[23] Raghavendra Ramachandra and Christoph Busch, "Presentation attack detection methods for face recognition systems: A comprehensive survey," *ACM Comput. Surv.*, vol. 50, no. 1, Mar. 2017.

[24] Davis E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[25] S. Milborrow and F. Nicolls, "Active Shape Models with SIFT Descriptors and MARS," *VISAPP*, 2014.

[26] S. Venkatesh, H. Zhang, R. Ramachandra, K. Raja, N. Damer, and C. Busch, "Can gan generated morphs threaten face recognition systems equally as landmark based morphs? - vulnerability and detection," in *2020 8th International Workshop on Biometrics and Forensics (IWBF)*, 2020, pp. 1–6.

[27] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4396–4405.

[28] Chris McCool, Sébastien Marcel, Abdenour Hadid, Matti Pietikäinen, Pavel Matějka, Jan Černocký, Norman Poh, Josef Kittler, Anthony Larcher, Christophe Lévy, Driss Matrouf, Jean-François Bonastre, Phil Tresadern, and Timothy Cootes, "Bi-modal person recognition on a mobile phone: using mobile phone data," in *IEEE ICME Workshop on Hot Topic in Mobile Multimedia*, 2012.

[29] 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, R. Ramachandra, and C. Busch, "Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting," in *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2017, pp. 1–7.

[30] A. Mohammadi, S. Bhattacharjee, and S. Marcel, "Deeply vulnerable: a study of the robustness of face recognition to presentation attacks," *IET Biometrics*, vol. 7, no. 1, pp. 15–26, 2018.