Towards Better Morphed Face Images Without Ghosting Artifacts

Clemens Seibold[1][a], Anna Hilsmann[1][b] and Peter Eisert[1,2][c]
[1]*Fraunhofer HHI, Berlin, Germany*
[2]*Humboldt University of Berlin, Berlin, Germany*

Keywords: Face Morphing Attacks, Ghosting Artifact Prevention, Morphed Face Images Dataset.

Abstract: Automatic generation of morphed face images often produces ghosting artifacts due to poorly aligned structures in the input images. Manual processing can mitigate these artifacts. However, this is not feasible for the generation of large datasets, which are required for training and evaluating robust morphing attack detectors. In this paper, we propose a method for automatic prevention of ghosting artifacts based on a pixel-wise alignment during morph generation. We evaluate our proposed method on state-of-the-art detectors and show that our morphs are harder to detect, particularly, when combined with style-transfer-based improvement of low-level image characteristics. Furthermore, we show that our approach does not impair the biometric quality, which is essential for high quality morphs.

1 INTRODUCTION

A morphed face image is a composite image that is generated by blending facial images of different subjects. Since the feasibility of tricking a facial recognition system to match two random subjects with one morphed face image was demonstrated by (Ferrara et al., 2014), a significant amount of research has been conducted in generating and detecting such images. Early publications on Morphing Attack Detection (MAD) relied on manually generated morphed face images for training and evaluation. However, since manual generation is a time-consuming task, automatic approaches paved the way for developing data-demanding machine learning-based detectors and evaluations on large datasets.

Most automatic face morphing approaches estimate the positions of facial landmarks in both input images (Makrushin et al., 2017), warp the images such that the landmarks have the same shape and position and then additively blend them. Images generated by these methods often suffer from ghosting artifacts caused by inaccuracies in the landmark position estimation or unalignable facial structures. These artifacts occur when two structures, such as the iris border, are not perfectly aligned. Figure 1 shows an example of a ghosting artifact. In the simple morphed face image, a second translucent border of the iris is visible due to inaccurate alignment of the iris shape. Our proposed method, however, prevents the appearance of such artifacts.

An alternative to the key-point-based method is the use of Generative Adversarial Networks (GANs) (Zhang et al., 2021). Images generated using GANs do not contain ghosting artifacts, but come with other limitations. For example, the resolution of these images is determined by the GAN architecture, the generation process is hard to control, and they often leave GAN-typical artifacts that allow their detection (Zhang et al., 2019).

In this paper, we address the prevention of ghosting artifacts in automatic key-point-based generation of morphed face images. In real attacks, attackers may manually correct the images to avoid ghosting artifacts, but this approach is impractical for large datasets. Unlike GAN-based methods, our proposed method allows results of any resolution. It uses a pixel-wise alignment technique that maps similar structures, such as the contour of the nostrils, iris, specular highlights etc. such that they have the same shape and position in both input images. Thus, it prevents ghosting artifacts in the final morphed face image. Figure 1 shows an example of a morphed face image generated using a simple key-point-based approach, our proposed improvement method, and a GAN-based approach.

[a] https://orcid.org/0000-0002-9318-5934
[b] https://orcid.org/0000-0002-2086-0951
[c] https://orcid.org/0000-0001-8378-4805

Seibold, C., Hilsmann, A. and Eisert, P.
Towards Better Morphed Face Images Without Ghosting Artifacts.
DOI: 10.5220/0012302800003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 3: VISAPP*, pages 272-281
ISBN: 978-989-758-679-8; ISSN: 2184-4321

(a) Simple Morph (b) Proposed Improvement (c) GAN-based Morph (d) Input Images

Figure 1: Morphed face image generated with a simple keypoint-based method (a), improved with the proposed ghosting artifact prevention method (b), generated with a GAN-based method (c), and input images (d). The ghosting artifact only appears in the simple approach. The GAN-based morph suffers from different artifacts, e.g., the unusual iris and pupil shapes.

We evaluate the impact of this enhancement on the detection rate of different state-of-the-art single image-based MAD approaches using uncompressed and compressed images, since studies showed that compressed images are harder to detect and differ strongly in feature space (Seibold et al., 2019b). Furthermore, we analyze its effect on the objective of face morphing attacks: Creating a face image that looks similar to two different subjects.

To evaluate our approach, we address the following four research questions:

R1: Does our proposed method make morphs harder to detect for single-image MAD techniques?

R2: Can detectors adapt to these novel morphed face images?

R3: Do the improved morphed face images still impersonate two different subjects?

R4: Does our proposed method still affect the detection rate of MAD techniques in compressed images?

In summary, our contributions are:

- A method to prevent ghosting artifacts in morphed face images as an additional component for keypoint-based morphing pipelines.
- A novel dataset of faultless morphed face images[1].
- An evaluation of different morphing methods on state-of-the-art detectors.

[1]Accessible under https://cvg.hhi.fraunhofer.de

The paper is structured as follows. The next section describes our pixel-wise improvement method. The experiments, including a short description of the used detectors and datasets, are presented in Section 3. Section 4 provides results of selected MAD techniques and the evaluation of the biometric quality of the used morphed face images.

2 RELATED WORK

Early research on morphing attack generation to study the feasibility of this attack and the detection of such images relied on manually generated morphed face images (Ferrara et al., 2014; Ramachandra et al., 2016). (Makrushin et al., 2017) proposed an automatic face morphing pipeline to generate visually faultless morphed face images, pushing the automatic generation of large data sets of morphed face images for the development of data-driven detection methods and their evaluation on large datasets. Several researchers adopted this concept and trained and evaluated their morphing detection methods on automatically generated morphed face images. However, only a very few authors have published their morphed face images or code for the generation of such. See (Hamza et al., 2022) for an overview of the generation and detection of morphed face images.

The mandatory blending process in face morphing pipelines usually impairs the quality of the images, often dampening the high spatial-frequency details, and causing the blended images to appear more dull than

the input images. (Seibold et al., 2021) proposed a method based on style transfer to counter this effect and showed that their improved attacks are harder to detect.

Other approaches to generate morphed face images are based on GANs. The first approaches were only capable of generating images in small resolutions, such as 64×64 pixels (Damer et al., 2018). Later approaches benefited from advances in GAN-based image generation, and (Zhang et al., 2021) proposed a face morphing method based on StyleGAN2 (Karras et al., 2020), which can create realistic face images in a resolution of 1024×1024 pixels.

3 PIXEL-WISE ALIGNMENT FOR MORPHED FACE IMAGE GENERATION

A typical face morphing pipeline consists of three main components: key-point-based alignment, additive blending, and an optional post-processing step to handle the background. We adopt the approach of (Seibold et al., 2017) to manage the background. This method involves copying the face of the morphed image into the background of one aligned input image with a smooth transition for the low spatial frequency components of the image and a sharp cut for the high spatial frequency part of the image between these two images. Ghosting artifacts occur when structures in the input images are not properly aligned, e.g. the nostrils have a different shape. Our approach can be seamlessly integrated into the morphing pipeline after the key-point-based alignment and before the additive blending.

3.1 Problem Formulation and Optimization

Pixel-wise alignment tasks are classically solved using the concept of the brightness constancy assumption (Horn and Schunck, 1981). Techniques based on this assumption aim to find a pixel warping from one image to another, minimizing the intensity difference between the warped and the target image. Likewise, we are looking for a warping that maps similar structures to the exactly same shape and same location, but focuses on characteristic structures to estimate the warping, e.g. borders of facial features such as specular highlights or the iris, instead of operating on intensity differences. Directly minimizing the intensity differences would lead to even worse aligned faces due to different skin tones or brightness variation. Thus, we first apply a spatially high-pass filter and only retain the high frequency information for our warping calculation.

We calculate two independent warping functions to warp each image I_1 and I_2 independently to intermediate aligned images. With $I(\mathbf{p})$ being the pixel intensity of an image at a pixel position $\mathbf{p} \in \mathcal{N}^2$, the warped image $\tilde{I}$ can be defined as

$$\tilde{I}(\mathbf{p}) = I(\mathbf{p} + w(\mathbf{p};\theta)), \tag{1}$$

with θ being the warp parameters, i.e. the $x-$ and $y-$ offsets per pixel.

The loss function for the data term of the alignment is

$$\begin{aligned}\mathcal{L}_d = (\theta_1,\theta_2) \sum_{\mathbf{p}\in\mathcal{P}} |I_1(\mathbf{p} + w(\mathbf{p};\theta_1)) \\ - I_2(\mathbf{p} + w(\mathbf{p};\theta_2))|_2^2,\end{aligned} \tag{2}$$

with $\mathcal{P}$ being the set of all pixel positions in the images. As this is an ill-posed problem, we add additional regularization terms that penalize the offset difference of neighboring pixels.

$$\mathcal{L}_s(\theta) = \sum_{(\mathbf{p_1},\mathbf{p_2})\in\mathcal{P}_n} |w(\mathbf{p_1};\theta) - w(\mathbf{p_2};\theta)|_2^2 \tag{3}$$

with $\mathcal{P}_n$ being neighboring pixels pairs, such that the second pixel is right or below the first pixel.

$$\mathcal{L}_b(\theta) = \sum_{\mathbf{p}\in\mathcal{P}_b} |w(\mathbf{p};\theta)|_2^2 \tag{4}$$

with $\mathcal{P}_n$ being the pixels at the border of the image or region of interest that is optimized.

The cost function to be minimized can thus be written as

$$\begin{aligned}\mathcal{L}(\theta_1,\theta_2) = \mathcal{L}_d(\theta_1,\theta_2) + \lambda\mathcal{L}_s(\theta_1) + \lambda\mathcal{L}_s(\theta_2) \\ + \lambda\mathcal{L}_b(\theta_1) + \lambda\mathcal{L}_b(\theta_2),\end{aligned} \tag{5}$$

with λ being a weighting factor for the smoothness term.

We minimize equation (5) using a Gauß-Newton algorithm. Minimizing equation (5) is a non-linear optimization problem, since the data term, in particular, $I_1(\mathbf{p})$ and $I_2(\mathbf{p})$ are usually non-linear. However, since the images are already pre-aligned, a large warp is not expected and $I_1(\mathbf{p})$ and $I_2(\mathbf{p})$ are assumed to behave partly linear for small changes. During each iteration we thus minimize the following system

$$\min_{w_{1,x},w_{1,y},w_{1,x},w_{2,y}} \left\| \mathbf{A}\cdot\mathbf{w} - \begin{bmatrix}\mathbf{i_2}-\mathbf{i_1}\\ \mathbf{0}\end{bmatrix} \right\|, \tag{6}$$

$$\text{with } A = \begin{bmatrix} G_{1,x} & G_{1,y} & -G_{2,x} & -G_{2,y} \\ \mathbf{P} & & & \\ & \mathbf{P} & & \\ & & \mathbf{P} & \\ & & & \mathbf{P} \end{bmatrix} \tag{7}$$

$$\text{and } \mathbf{w} = \begin{bmatrix}\mathbf{w}_{1,x}^T & \mathbf{w}_{1,y}^T & \mathbf{w}_{2,x}^T & \mathbf{w}_{2,y}^T\end{bmatrix}^T \tag{8}$$

and $\mathbf{i_n}$ being the vectorized images I_n, $G_{n,x}/G_{n,x}$ diagonal matrices that contain the image gradient of I_n in x-/y direction, $\mathbf{w}_{n,x}/\mathbf{w}_{n,y}$ the pixel motion in x-/y-direction and P a sparse matrix that describes the smoothness term as defined in Equations (3) and (4) scaled by $\sqrt{\lambda}$. It contains for every unordered pair of neighboring pixel one sparse row with $\sqrt{\lambda}$ at the column that represents the left or upper pixel and $-\sqrt{\lambda}$ at the column that represents the other pixel. For every border pixel, there is one row with only a non-zero entry in the column that represents that pixel, with a value of $\sqrt{\lambda}$. The optimal solution to this problem can be obtained by solving

$$A^T A\mathbf{w} = A^T[\mathbf{i_2} - \mathbf{i_1}\ \mathbf{0}]^T. \quad (9)$$

The matrix $A^T A$ is sparse but very large. Instead of explicitly setting it up, we utilize the Minimal Residual (MINRES) method to numerically solve equation (9) (Paige and Saunders, 1975). The MINRES method tackles the minimization problem through an iterative approach, requiring only a procedure for right-multiplication of arbitrary vectors $\mathbf{x}$ with the matrices A and A^T. If we treat the vector $\mathbf{x}$ as an image, the multiplications related to the data term can be performed through pixel-wise operations with the image gradients. Similarly, the smoothness term can be implemented using convolution techniques.

3.2 Examples of Improved Morphed Face Images

Figure 2 shows further examples of our proposed method. The first example demonstrates the effectiveness of our method to avoid a morphing artifact around the nostrils. This particular ghosting artifact commonly occurs with automatic face morphing pipelines, since the upper part of the nostrils is not estimated by standard facial landmark detector such as (Kazemi and Sullivan, 2014), which is often used due to its availability in DLib (King, 2009). Another often arising ghosting artifact is caused by misaligned specular highlights in the eyes as shown the second example. Again, these artifacts are avoided by the proposed method.

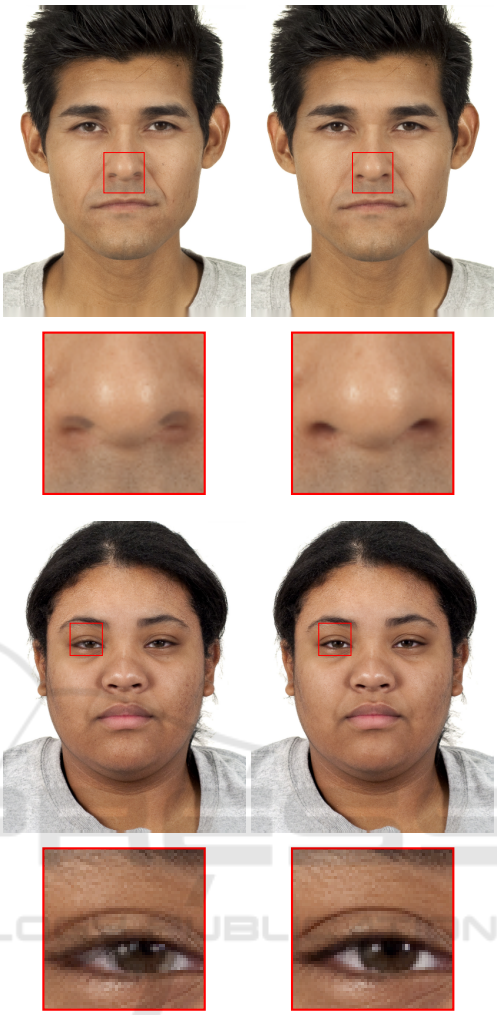

Figure 2: Examples of different ghosting artifacts for a simple morphed face image (left) and our improved approach (right). The artifacts are avoided by the proposed alignment method.

4 EXPERIMENTAL SETUP

4.1 Datasets

For the training and evaluation of detectors, we compiled a large dataset of bona fide images from various sources and generated morphed face images using different methods, as described below.

We collected images from publicly available datasets, including BU-4DFE (Zhang et al., 2014), CFD (Ma et al., 2015), CFD-India (Lakshmi et al., 2020), CFD-MR (Ma et al., 2020), FERET (Phillips et al., 1998), MR2 (Strohminger N, Gray K, Chituc V, Heffner J, Schein C, Heagins TB, 2016), FRLL (DeBruine and Jones, 2021), PUT (Kasiński et al., 2008), scFrontal (Grgic et al., 2011), SiblingDB (Vieira et al., 2014), Utrecht[2], YAWF (DeBruine and Jones, 2017), RADIATE (Conley et al., 2018), Ecua (Avilés et al., 2019), CUFS (Wang and Tang, 2009), Iranian Women[2], AMFD (Chen et al., 2021), stir[2], FED (Aifanti et al., 2010) and FRGCv2 (Phillips et al., 2005). Additionally, we used in-house datasets and acquired additional face images through search

[2]https://pics.stir.ac.uk/

engines. All images underwent manual checks to ensure that the subjects were in a neutral pose, looking directly into the camera, free from occlusions, and that a minimum inter-eye distance of 90 pixels was maintained. Furthermore, each subject was included only once in the dataset.

The FRGCv2 and FRLL datasets were exclusively selected for testing purposes. The other datasets, referred to as the mixed dataset, were divided into a training set with 70% of all images and a testing and validation set with 15% each. The mixed dataset consists of about 9,200 images with 6,400 images used for training, 1,440 for testing, and 1,400 for validation. The FRLL has 102 bona fide image with a neutral pose. From the FRGCv2 set, we utilized about 1,441 uniformly illuminated images with a neutral head pose and a uniform background for morph generation and further 1,726 images as reference images for the evaluation of the attack success on facial recognition systems.

Before using the images for training and evaluation, the faces were cropped such that they show the head and parts of the shoulder, as recommended by the ICAO (International Civil Aviation Organization, 2018) for facial images stored on passports. After cropping, the images were resized to 513x431, which is a common size for passports (Neubert et al., 2018).

4.2 Morphed Face Images Generation

We created morphed face images using five different methods or combinations of methods. The simple morphs were generated using the pipeline from (Seibold et al., 2020). The *ST* morphs are an improved version of the simple morphs, incorporating the style-transfer-based improvement described in (Seibold et al., 2019a). We refer to the simple morphs improved with our proposed pixel-wise alignment method as *PW* morphs. When using both of the methods we refer to them as *PWST* morphs. For creating GAN-based morphs, we use the method of (Zhang et al., 2021) and refer to them as *MIP2* morphs. To generate the *MIP2* morphs, we used the implementation of (Sarkar et al., 2022).

To select suitable pairs for generating morphed face images, we followed the protocol in (Scherhag et al., 2020) for the FRGCv2 dataset and in (Neubert et al., 2018) for the FRLL dataset. For the generation of morphed face images from the other dataset, we selected the pair such that they are both from the same dataset and their gender and ethnicity match. The number of morphs and bona fide images in the respective set is the same. During testing and validation, the data is augmented by horizontal flipping. The validation set is specifically used for the evaluation of the epochs of detectors based on Deep Neural Networks (DNNs), selecting the best performing model.

4.3 Detectors

We investigate the impact of our proposed ghosting artifact prevention method on the detection rates of five detectors. One of the detectors is based on an ensemble of features and utilizes a probabilistic CRC for classification (Ramachandra et al., 2019). The remaining detectors employ DNNs (Seibold et al., 2021). All methods only operate on the inner part of the face as proposed by their authors.

The DNN-based detectors in our study use one output neuron and are all trained using a binary cross-entropy loss. The detector *VGG-A naïve* uses the VGG-A architecture and *Xception* the Xception architecture, which has demonstrated effectiveness in detecting Deep Fakes (Malolan et al., 2020). The *Feature Focus* (Seibold et al., 2021) detector incorporates an additional loss that activates half of the neurons in the last convolutional layer strongly for morphed face images, and the other half for bona fide images. Inspired by the work of (Ramachandra et al., 2019) on effects of color spaces on the performance of morphing detectors, we tested the most promising detector on images in the HSV color space. These are denoted with (HSV). The Feature Focus detector in HSV color space, trained only randomly compressed images, was also submitted to the Face Morphing Detection challenge of the NIST. It showed an outstanding performance and took first place in different categories (Ngan et al., 2023).

The *Feature Ensemble* detector (Venkatesh et al., 2020) splits the images into two different color spaces and calculates a Laplacian pyramid with three levels. For each of the resulting images, a Histogram of Gradients, Binarized Statistical Image Features and Local Binary Pattern are calculated and a probabilistic CRC classifier is employed.

5 RESULTS

In the following, each subsection will answer one of the research questions presented in the Introduction.

5.1 Are Our Improved Morphs Harder to Detect?

To examine the impact of the ghosting artifact removal and address **R1**, we trained all detectors on

Table 1: EER and BPCER@APCER=5% for training on simple morphs only to analyze the effect of different morphed face image improvement methods and a GAN-based generation method. The best performing attack method is highlight in bold and the second best are underlined. The PWST morphing method, which is our proposed method in combination with an improvement based on style transfer (Seibold et al., 2019a), achieves in all cases the highest or second highest error rates and thus the improved morphed face images are harder to detect than the simple morphed face images without any improvement applied.

Morph Method Dataset	Equal-Error-Rates[%] simple	PW (ours)	ST	PWST (ours+ST)	MIP2	BPCER[%]@APCER=5% simple	PW (ours)	ST	PWST (ours+ST)	MIP2
Detector: *Feature Ensemble* (Venkatesh et al., 2020)										
Mixed Set	5.03	5.73	20.67	**21.68**	15.24	5.24	6.56	49.51	**49.83**	37.72
FRLL	1.96	1.96	10.29	11.27	**12.75**	0.00	0.49	15.69	17.16	**25.00**
FRGCv2	3.46	4.93	14.13	15.96	**22.36**	2.03	4.78	29.67	36.69	**60.92**
Detector: *VGG-A* (Simonyan and Zisserman, 2015)										
Mixed Set	0.59	1.67	11.61	16.50	**22.22**	0.14	0.73	21.01	30.87	**78.33**
FRLL	0.00	0.00	0.49	1.96	**13.24**	0.00	0.00	0.00	0.49	**61.76**
FRGCv2	0.21	1.17	2.10	6.46	**9.45**	0.00	0.25	1.37	7.62	**17.99**
Detector: *Xception* (Chollet, 2017)										
Mixed Set	0.38	1.39	6.78	10.70	**11.01**	0.07	0.24	8.65	18.30	**30.10**
FRLL	0.05	0.71	3.43	7.84	**12.75**	0.00	0.00	1.96	12.25	**43.63**
FRGCv2	0.47	2.34	6.41	**12.45**	10.53	0.05	1.32	7.77	23.07	**24.24**
Detector: *Feature Focus (RGB)* (Seibold et al., 2021)										
Mixed Set	0.73	1.29	11.64	**15.11**	13.40	0.07	0.63	21.63	29.65	**51.15**
FRLL	0.00	0.00	0.49	1.96	**10.31**	0.00	0.00	0.00	1.96	**29.90**
FRGCv2	0.21	0.46	1.84	6.05	**7.99**	0.00	0.20	0.71	7.27	**20.58**
Detector: *Feature Focus (HSV)* (Seibold et al., 2021)										
Mixed Set	0.97	1.04	12.23	**13.45**	7.15	0.07	0.14	21.84	**26.88**	12.33
FRLL	0.00	0.00	1.03	1.56	**2.45**	0.00	0.00	0.00	0.00	**0.49**
FRGCv2	0.00	0.00	4.67	**6.96**	5.55	0.00	0.00	4.62	**9.71**	7.06

simple: (Seibold et al., 2020); ST: (Seibold et al., 2019a); MIP2: (Zhang et al., 2021)

simple morphs and evaluated them on all types of morphs. We report the detectors' performance using Attack Presentation Classification Error Rates (APCER) and Bona fide Presentation Classification Error Rates (BPCER) as defined in ISO/IEC 30107-3 (International Organization for Standardization, 2017) and Equal Error Rates (EER). The BPCER is reported at a fixed APCER of 5%. Table 1 reveals that the removal of ghosting artifacts has only a small impact on the detection, in contrast to the *ST* improvement or the utilization of GANs for morph generation. However, these morphs are still harder or at least as hard to detect. While the difference in the EERs for the simple morphs and the *PW* morphs is always smaller than 2%, the EERs for the *ST* morphs are more than 10% larger and the EER for the *MIP2* morphs is even up to 20% larger for the *VGG-A naïve* detector. In combination with the style-transfer-based improvement, however, the error rates notably increase compared to using style-transfer only for the improvement.

5.2 Can the Detectors Adapt to the New Challenge?

To assess whether the detectors can adapt to the proposed improved morphs and the other types of morphs (**R2**), we added *PWST* and *MIP2* morphs to the training data. The results are shown in Table 2. For the DNN-based detectors, the error rates significantly decreased for the *PWST* and *MIP2* morphs in nearly all cases. The error rates for the PW and ST morphs drop in most cases, but the rates for the simple morphs slightly increase in most cases. The *Feature Ensemble* detector shows the largest error rates and has also the strongest increase in error rates for the simple morphs. To answer **R2**: The DNN-based detectors can easily adapt to the improved and to the *MIP2* morphs by just adding examples of these morphs to the training data. The *Feature Focus (HSV)* detector shows the best performance.

Table 2: EER and BPCER@APCER=5% for training on simple, *PWST* and *MIP2* morphs to analyze if the detectors can adapt to the threat of improved and GAN-based morphed face images. The best-performing attack method is highlight in bold and the second best are underlined. The DNN-based detectors seem to be able to learn other traces of forgery to distinguish between bona fide and morphed face images.

Morph Method Dataset	Equal-Error-Rates[%] simple	PW (ours)	ST	PWST (ours+ST)	MIP2	BPCER[%]@APCER=5% simple	PW (ours)	ST	PWST (ours+ST)	MIP2
Detector: *Feature Ensemble* (Venkatesh et al., 2020)										
Mixed Set	14.21	13.97	14.52	**14.87**	4.34	**29.27**	28.72	28.30	28.72	4.06
FRLL	**7.84**	**7.84**	5.88	5.98	0.53	**10.78**	**10.78**	6.86	6.86	0.00
FRGCv2	10.06	11.00	13.72	**14.48**	2.59	17.89	22.21	28.81	**32.72**	1.27
Detector: *VGG-A* (Simonyan and Zisserman, 2015)										
Mixed Set	1.15	0.69	**2.19**	1.74	0.87	0.28	0.17	**1.01**	0.59	0.14
FRLL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FRGCv2	1.95	2.08	**5.71**	5.29	1.14	0.56	0.81	**6.30**	5.44	0.15
Detector: *Xception* (Chollet, 2017)										
Mixed Set	0.56	0.31	**1.25**	0.90	0.52	0.03	0.00	**0.21**	0.14	0.00
FRLL	0.00	0.00	0.07	0.02	**0.49**	0.00	0.00	0.00	0.00	0.00
FRGCv2	1.37	1.84	5.13	**5.24**	0.41	0.41	0.66	5.28	**5.64**	0.00
Detector: *Feature Focus (RGB)* (Seibold et al., 2021)										
Mixed Set	1.20	1.84	3.50	**4.61**	0.83	0.28	0.92	2.44	**4.05**	0.14
FRLL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FRGCv2	1.69	2.70	5.39	**6.77**	0.83	0.71	1.88	5.89	**8.59**	0.15
Detector: *Feature Focus (HSV)* (Seibold et al., 2021)										
Mixed Set	1.04	0.73	1.15	**1.18**	0.49	0.28	0.21	0.31	**0.38**	0.07
FRLL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FRGCv2	0.00	0.00	**0.15**	0.10	0.00	0.00	0.00	0.00	0.00	0.00

simple: (Seibold et al., 2020); ST: (Seibold et al., 2019a); MIP2: (Zhang et al., 2021)

Table 3: MMPMR@FAR0.1% for FRGCv2 dataset to analyse if the morphed face images portray two different subjects. The improvement methods seem to have only a marginal impact on the biometric properties of the face images and after applying the improvement methods (ours + ST), the attacks are more successful than the baseline (simple). Whether the biometric quality of the GAN-based morphs (MIP2) is better or much worse than these of the keypoint-based morphs strongly depends on the facial recognition system used for the evaluation.

	Bona Fide FRR	FAR	MMPMR@FAR0.1% simple (Seibold et al., 2020)	PWST (ours + ST)	MIP2 (Zhang et al., 2021)
ArcFace	1.18%	0.1%	31.48%	32.57%	**37.97**%
COTS	0.00%	0.02%	48.50%	**48.65**%	17.63%

5.3 Do the Improved Morphs Still Impersonate Two Different Subjects?

To address **R3**, we evaluated the biometric quality using the MinMax-Mated Morph Presentation Match Rate (MMPMR) (Scherhag et al., 2017) on the FRGCv2 dataset using the protocol of (Scherhag et al., 2020). Two different facial recognition systems were employed for the evalution: An implementation of ArcFace[3] and a commercial off-the-shelf (COTS) system. The false acceptance rate (FAR) threshold for the COTS system was determined based on its manual and for the ArcFace system calculated from the FRGCv2 dataset. Table 3 shows that the improvement methods have only a minor effect on the biometric quality and they even improve the success rate of the attacks slightly. Another interesting finding is that the *MIP2* morphs are much better at tricking the ArcFace system than the COTS system and that they perform much worse on the COTS than the other morphs do.

[3]https://github.com/mobilesec/arcface-tensorflowlite

Table 4: Error rates of the Feature Focus (HSV) detector for compressed images. The best performing attack method is highlight in bold and the second best are underlined. Also for the JPG-compressed images, the *PWST* and MIP2 morphs are much harder to detect than the simple morphs. The detector can adapt to the improved morphs, but does still perform much worse than on the simple attacks.

Morph	Equal-Error-Rates[%]					BPCER[%]@APCER=5%				
Method	simple	PW	ST	PWST	MIP2	simple	PW	ST	PWST	MIP2
Dataset		(ours)		(ours+ST)			(ours)		(ours+ST)	
	Detector: *Festure Focus (HSV)* (Seibold et al., 2021) train on simple only									
Mixed Set	3.41	6.60	17.93	<u>24.67</u>	**25.14**	2.12	8.54	37.67	<u>55.17</u>	**76.63**
FRLL	2.60	5.59	14.22	<u>21.57</u>	**32.36**	0.00	6.37	22.55	<u>48.53</u>	**79.41**
FRGCv2	1.02	3.76	11.85	<u>20.38</u>	**21.80**	0.15	3.00	22.26	<u>46.19</u>	**78.81**
Method	simple	PW	ST	PWST	MIP2	simple	PW	ST	PWST	MIP2
Dataset		(ours)		(ours+ST)			(ours)		(ours+ST)	
	Detector: *Festure Focus (HSV)* (Seibold et al., 2021) train on simple, PWST and MIP2									
Mixed Set	4.83	5.11	<u>10.46</u>	**11.19**	6.46	4.65	5.21	<u>18.61</u>	**18.72**	7.74
FRLL	3.93	6.37	<u>10.78</u>	**14.78**	11.42	3.43	7.84	<u>17.65</u>	**26.47**	21.57
FRGCv2	2.80	3.30	**7.67**	<u>7.62</u>	7.37	1.22	1.73	<u>10.82</u>	**11.48**	9.96

simple: (Seibold et al., 2020); ST: (Seibold et al., 2019a); MIP2: (Zhang et al., 2021)

5.4 Do the Improvements Make a Difference for JPG-Compressed Images?

Table 4 shows the error rates of the Feature Focus (HSV) detector trained and tested on compressed images to answer **R4**. We used a compression rate that targets a file size between 15kB and 20kB, which is the typically mandatory and reserved size for storing facial image on the passport chips (International Civil Aviation Organization, 2015). The error rates for the morphs improved by our ghosting artifact prevention method are larger than those for the simple morphs. In combination with the style-transfer improvement, the error rates are even larger compared to when only one of them is used. Thus, our proposed ghosting artifact prevention also affects the detection rates in compression face images.

6 SUMMARY AND DISCUSSION

In this paper, we introduced a ghosting artifact prevention method that can be integrated into key-point based face morphing pipelines. The prevention of ghosting artifacts can be performed manually by an attacker, but this is not feasible for large training or evaluation datasets. Our approach effectively prevents ghosting artifacts without compromising the biometric quality of the morphs. Furthermore, it poses a greater challenge for MAD techniques to detect these improved morphs. In combination with the style-transfer-based improvement method of (Seibold et al., 2019a), the resulting morphed face images provide a new challenge for MAD techniques. One of its biggest advantages compared to GAN-based methods is that our method can produce morphs in any resolution, while the resolution of GAN-morphs is limited by the GAN's architecture. Furthermore, the keypoints-based approach allows a better control over the morphed face image generation process. This control includes factors such as balancing the influence of each individual input image on the resulting morphed face image (blending factor), specifying which regions should be blended, and other parameters. The keypoint-based morphing pipeline closely aligns with the approach an attacker might use to create a morphed face images, utilizing publicly-available tools for warping and blending. By incorporating the improved morphs into the training data, we observe enhanced detection performance for these improved morphs. This further highlights the effectiveness and practical significance of our approach in improving the detection capabilities of MAD techniques. In future work, we plan to study the impact of the our proposed improvement on detectors that analyze the shape of reflections, given that pixel-wise alignment changes the face's geometry (Seibold et al., 2018).

ACKNOWLEDGEMENTS

This work has recieved partial funding by the German Federal Ministry of Education and Research (BMBF) through the Research Program FAKEID under Contract no. 13N15735, as well as the Fraunhofer Society in the Max Planck-Fraunhofer collaboration project NeuroHum.

REFERENCES

Aifanti, N., Papachristou, C., and Delopoulos, A. (2010). The mug facial expression database. In *WIAMIS*. IEEE.

Avilés, J., Toapanta, H., Morillo, P., and Vallejo-Huanga, D. (2019). Dataset of Ethnic Facial Images of Ecuadorian People.

Chen, J. M., Norman, J. B., and Nam, Y. (2021). Broadening the stimulus set: introducing the American multiracial faces database. *Behavior Research Methods*.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258.

Conley, M. I., Dellarco, D. V., Rubien-Thomas, E., Cohen, A. O., Cervera, A., Tottenham, B., and Casey, B. (2018). The racially diverse affective expression (radiate) face stimulus set. *Psychiatry Research*, 270:1059–1067.

Damer, N., Saladié, A. M., Braun, A., and Kuijper, A. (2018). 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)*, pages 1–10.

DeBruine, L. and Jones, B. (2017). Young Adult White Faces with Manipulated Versions. https://figshare.com/articles/dataset/Young_Adult_White_Faces_with_Manipulated_Versions/4220517.

DeBruine, L. and Jones, B. (2021). Face Research Lab London Set. https://figshare.com/articles/dataset/Face_Research_Lab_London_Set/5047666.

Ferrara, M., Franco, A., and Maltoni, D. (2014). The magic passport. In *IEEE International Joint Conference on Biometrics*.

Grgic, M., Delac, K., and Grgic, S. (2011). SCface — Surveillance Cameras Face Database. *Multimedia Tools and Applications*, 51(3):863–879.

Hamza, M., Tehsin, S., Humayun, M., Almufareh, M. F., and Alfayad, M. (2022). A comprehensive review of face morph generation and detection of fraudulent identities. *Applied Sciences*, 12(24).

Horn, B. K. and Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17(1):185–203.

International Civil Aviation Organization (2015). Doc 9303 - machine readable travel documents - part 10.

International Civil Aviation Organization (2018). Technical report - portrait quality (reference facial images for mrtd).

International Organization for Standardization (2017). Iso/iec 30107-3:2017 information technology – biometric presentation attack detection – part 3: Testing and reporting.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., and Aila, T. (2020). Analyzing and improving the image quality of stylegan. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8107–8116.

Kasiński, A., Florek, A., and Schmidt, A. (2008). The PUT face database. *Image Processing and Communications*, 13:59–64.

Kazemi, V. and Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10(60):1755–1758.

Lakshmi, Wittenbrink, Correll, and Ma (2020). The India Face Set: International and Cultural Boundaries Impact Face Impressions and Perceptions of Category Membership. *Frontiers in Psychology*.

Ma, Kantner, and Wittenbrink (2020). Chicago Face Database: Multiracial Expansion. *Behavior Research Methods*.

Ma, D., Correll, J., and Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47.

Makrushin, A., Neubert, T., and Dittmann, J. (2017). Automatic generation and detection of visually faultless facial morphs. In *VISIGRAPP*, pages 39–50. SciTePress.

Malolan, B., Parekh, A., and Kazi, F. (2020). Explainable deep-fake detection using visual interpretability methods. In *2020 3rd International Conference on Information and Computer Technologies (ICICT)*, pages 289–293.

Neubert, T., Makrushin, A., Hildebrandt, M., Kraetzer, C., and Dittmann, J. (2018). Extended stirtrace benchmarking of biometric and forensic qualities of morphed face images. *IET Biometrics*, 7(4):325–332.

Ngan, M., Patrick, G., Hanaoka, K., and Kuo, J. (2023). Face recognition vendor test (frvt) part 4: Morph - performance of automated face morph detection.

Paige, C. C. and Saunders, M. A. (1975). Solution of sparse indefinite systems of linear equations. *SIAM Journal on Numerical Analysis*, 12(4):617–629.

Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., and Worek, W. (2005). Overview of the face recognition grand challenge. In *CVPR'05*, volume 1, pages 947–954 vol. 1.

Phillips, P., Wechsler, H., Huang, J., and Rauss, P. J. (1998). The feret database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306.

Ramachandra, R., Raja, K., and Busch, C. (2016). Detecting morphed face images. In *Proc. International Conference on Biometrics Theory, Applications and Systems (BTAS)*.

Ramachandra, R., Venkatesh, S., Raja, K., and Busch, C. (2019). Towards making morphing attack detection robust using hybrid scale-space colour texture features. In *IEEE 5th International Conference on Identity, Security, and Behavior Analysis*, pages 1–8.

Sarkar, E., Korshunov, P., Colbois, L., and Marcel, S. (2022). Are gan-based morphs threatening face recognition? In *ICASSP 2022 - 2022 IEEE International*

Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 2959–2963.

Scherhag, U., Nautsch, A., Rathgeb, C., Gomez-Barrero, M., Veldhuis, R. N. J., Spreeuwers, L., Schils, M., Maltoni, D., Grother, P., Marcel, S., Breithaupt, R., Ramachandra, R., and Busch, C. (2017). Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting. In *BIOSIG*, pages 1–7.

Scherhag, U., Rathgeb, C., Merkle, J., and Busch, C. (2020). Deep face representations for differential morphing attack detection. *IEEE Transactions on Information Forensics and Security*, 15:3625–3639.

Seibold, C., Hilsmann, A., and Eisert, P. (2018). Reflection analysis for face morphing attack detection. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 1022–1026.

Seibold, C., Hilsmann, A., and Eisert, P. (2019a). Style your face morph and improve your face morphing attack detector. In *BIOSIG*, pages 35–45.

Seibold, C., Hilsmann, A., and Eisert, P. (2021). Feature focus: Towards explainable and transparent deep face morphing attack detectors. *Computers*, 10(9).

Seibold, C., Hilsmann, A., Makrushin, A., Kraetzer, C., Neubert, T., Dittmann, J., and Eisert, P. (2019b). Visual feature space analyses of face morphing detectors. In *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6.

Seibold, C., Samek, W., Hilsmann, A., and Eisert, P. (2017). Detection of face morphing attacks by deep learning. In *Digital Forensics and Watermarking*, pages 107–120, Cham. Springer International Publishing.

Seibold, C., Samek, W., Hilsmann, A., and Eisert, P. (2020). Accurate and robust neural networks for face morphing attack detection. *JISA*, 53:102526.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Strohminger N, Gray K, Chituc V, Heffner J, Schein C, Heagins TB (2016). The mr2: A multi-racial, mega-resolution database of facial stimuli. *Behav Res Methods*.

Venkatesh, S., Ramachandra, R., Raja, K., and Busch, C. (2020). Single image face morphing attack detection using ensemble of features. In *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, pages 1–6.

Vieira, T. F., Bottino, A., Laurentini, A., and De Simone, M. (2014). Detecting siblings in image pairs. *The Visual Computer*, 30(12):1333–1345.

Wang, X. and Tang, X. (2009). Face Photo-Sketch Synthesis and Recognition. *PAMI*.

Zhang, H., Venkatesh, S., Ramachandra, R., Raja, K., Damer, N., and Busch, C. (2021). Mipgan -generating strong and high quality morphing attacks using identity prior driven gan. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, PP:1–1.

Zhang, X., Karaman, S., and Chang, S.-F. (2019). Detecting and simulating artifacts in gan fake images. In *WIFS*, pages 1–6.

Zhang, X., Yin, L., Cohn, J. F., Canavan, S. J., Reale, M., Horowitz, A., Liu, P., and Girard, J. M. (2014). Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. *Image Vis. Comput.*, 32(10):692–706.