3D FACE MORPH GENERATION USING GEOMETRY-AWARE TEMPLATE INVERSION

Hatef Otroshi Shahreza[1,2], *Laurent Colbois*[1,3], *Sébastien Marcel*[1,3]

[1]Idiap Research Institute, Martigny, Switzerland
[2]École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[3]Université de Lausanne (UNIL), Lausanne, Switzerland

{hatef.otroshi, laurent.colbois, sebastien.marcel}@idiap.ch

ABSTRACT

While face recognition systems have become a popular solution for applications which require automatic authentication, their vulnerability to morphing attacks has become a major concern in sensitive scenarios. This work proposes a novel method to generate 3D face morphs. Given two source images, we use their face embeddings to derive an optimal morph embedding, and then use a geometry-aware template inversion method based on Generative Neural Radiance Fields (GNeRF) to construct a 3D face morph from this optimal embedding. Leveraging from the GNeRF structure, we can generate morph images with any arbitrary view-point. Our experiments show that our method achieve comparable performance with previous morph generation methods from the literature, and has an additional advantage of generating 3D results. To our knowledge, this is the first work on generating 3D face morphs based on GNeRF models, and it can potentially be used for sophisticated morphing attacks. The source code of our experiments is publicly released.

Index Terms— 3D Morph, Embeddings, Face Recognition, Geometry-aware, Morph Attack, Neural Radiance Fields (NeRF), Template Inversion (TI)

1. INTRODUCTION

Face recognition (FR) systems are being widely deployed in numerous applications for identity recognition. In spite of their high accuracy, FR systems are shown to be vulnerable to different attacks, such as face morphing attacks. Morphing attacks aims at mixing the faces of two contributing subjects to generate a so-called *morph*, and enroll the morph image as reference into a FR system. In successful attacks, both contributing subjects can then be matched by the FR system to the same morph reference. This type of attack is of great concern in border control situations, given it enables multiple actors to share a single passport. As some countries (e.g., UAE [1]) are considering the introduction of 3D photos as passport picture reference, there is a new interest in studying the risk associated with 3D morphing attacks.

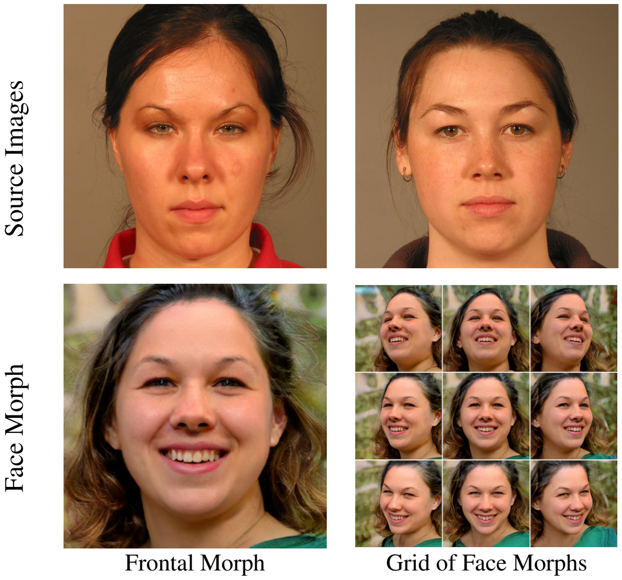

Fig. 1: Sample face morph gererated by our method.

Morphing attacks have initially been introduced in [2], which proposed to take two face images, align them across several face landmarks using affine warping, and average their pixels to produce an image practically representing a mixture of both identities (also known as landmark-based morphing). While the original method required manual work using image editing software, the process was automated and improved in subsequent works, e.g., [3].

With development of deep learning and image generation methods, a new broad category of morphing attacks has emerged, to generate morph image by interpolating face representations and decoding this interpolated representation into a morph. The seminal work in this field is MorGAN [4], which fully trains a system composed of an encoder of

This research is based upon work supported by the H2020 TReSPAsS-ETN Marie Skłodowska-Curie early training network (grant agreement 860813) and the Swiss Center for Biometrics Research & Testing at Idiap Research Institute.

979-8-3503-2411-2/25/$31.00 ©2025 IEEE

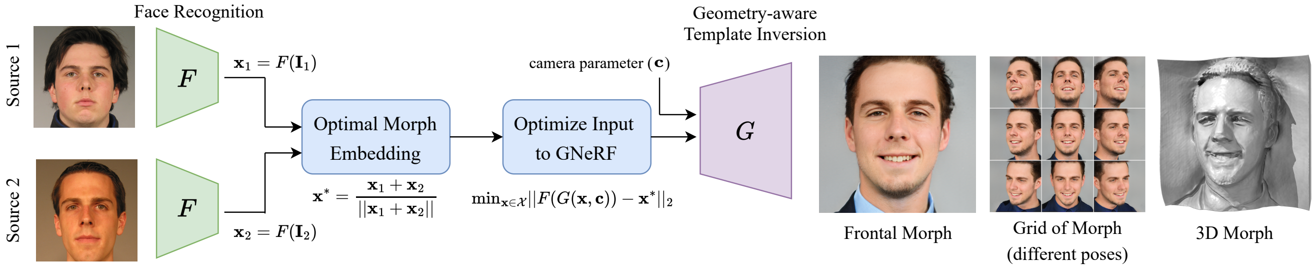
Face Recognition
Source 1
Source 2
F
F
$\mathbf{x}_1 = F(\mathbf{I}_1)$
$\mathbf{x}_2 = F(\mathbf{I}_2)$
Optimal Morph Embedding
$\mathbf{x}^* = \frac{\mathbf{x}_1 + \mathbf{x}_2}{||\mathbf{x}_1 + \mathbf{x}_2||}$
Optimize Input to GNeRF
$\min_{\mathbf{x} \in \mathcal{X}} ||F(G(\mathbf{x}, \mathbf{c})) - \mathbf{x}^*||_2$
camera parameter ($\mathbf{c}$)
Geometry-aware Template Inversion
G
Frontal Morph
Grid of Morph (different poses)
3D Morph

Fig. 2: Block-diagram of the proposed method.

face images into a latent space, and a generative adversarial network able to reconstruct those faces from their latent representation. Interpolation between faces can then be computed in the latent space. However, the generated morphs were of relatively low-resolution, in particular not ICAO compliant. Following works [5, 6] solve this issue by instead exploiting a pretrained StyleGAN network to generate high-resolution faces. The same idea of encoding-interpolation-generation applies, but in the absence of an encoder jointly trained with the model, the encoding step is performed through an optimization process, that finds latent representations of each source images by minimizing the perceptual loss between the real and resynthesized faces. The MIPGAN method [7] further builds on this idea by eliminating the interpolation step, and instead directly finding a good latent representation of the morph through the inclusion of a biometric loss in the optimization process. Later works [8] apply similar ideas using diffusion models instead as the generative backbone.

More recently, a new approach based on face template inversion has been explored. The core idea is that, given some model that can invert a face recognition network (i.e., reconstruct face images from face embeddings), the morphing can be directly performed arithmetically in the face embedding space (by computing the theoretical optimal morph embedding), and then a good candidate morph can be generated by inverting this optimal morph embedding. In other words, the encoding-interpolation-generation framework from previous works still applies, but the encoder is the face recognition network itself, the latent space is the face embedding space, and the decoder is the template inverter. In [9], authors assumed that the face embedding alone does not contain enough information to reconstruct a face image, and thus trained an ad-hoc additional autoencoder to extract all the identity-agnostic content from the source images. This encoded content is available to the decoder to help with face reconstruction. However, advances in template inversion research reveals this extra autoencoder is not necessary. This is showcased in [10], which generated morphs using template inversion methods [11, 12]. Template inversion methods keep evolving [13, 14], and this evolution directly translates into applications for morph generation by through inversion of optimal morph embeddings. Our work lies in the continuity of this line of research and we particularly focus on generating morph images using geometry-aware 3D face reconstruction, as presented in [14]. To our knowledge, this is the first work to generate 3D face morph based on Neural Radiance Fields (NeRF), which allows synthesizing face morph with any arbitrary view-point. A previous work [15] proposes a landmark-based 3D morphing attack, but relies on the source images themselves to be 3D (point cloud), which in particular makes it hard to compare with previous 2D morphing attacks. In contrast, our method enables to create a 3D morph from a set of 2D source images.

With this work, we want to highlight the strong link between template inversion research and morphing attack generation and detection research. As template inversion methods keep evolving, we believe important to keep reevaluating what danger they represent when used for morphing attack generation. We focus here specifically on the application of our method as a 2D attack (i.e., using a fixed camera view point). Our experiments show that resulting morphs cause enough FR vulnerability to be a concern (with an effectiveness comparable to previous 2D methods from the literature), but have the additional strengths of being 3-dimensional. Fig. 1 illustrates a sample morph image generated by our method.

2. METHODOLOGY

Let $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ denote two source images, $F(.)$ a facial feature extractor which extract face embeddings $\boldsymbol{x}_i = F(\boldsymbol{I}_i) \in \mathcal{X}$. Also, let $d(.,.)$ denote a distance metric on $\mathcal{X}$. Then, the optimal morph embedding can be defined by:

$$\boldsymbol{x}^* = \underset{\boldsymbol{x} \in \mathcal{X}}{\arg\min} \left[d(\boldsymbol{x}_1, \boldsymbol{x}) + d(\boldsymbol{x}_2, \boldsymbol{x}) \right]. \tag{1}$$

If we assume source embeddings to be normalized and the cosine distance as distance metric, the optimal morph embedding can be computed as follows:

$$\boldsymbol{x}^* = \frac{\boldsymbol{x}_1 + \boldsymbol{x}_2}{||\boldsymbol{x}_1 + \boldsymbol{x}_2||} \tag{2}$$

An ideal morph generation algorithm would generate face images whose embeddings are exactly optimal morph embeddings. To this end, we use a state-of-the-art template inversion method to generate face morphs from optimal morph embeddings. We use the geometry-aware face reconstruction (GaFaR) method proposed in [14], which is based on a generative NeRF model that can generate face images with any arbitrary view (pose). The GaFaR model $G(.,.)$ takes the face embedding $\boldsymbol{x}$ and camera parameter $\boldsymbol{c}$ as input and generate face image $\hat{\boldsymbol{I}} = G(\boldsymbol{x}, \boldsymbol{c})$. Therefore, we can generate the morph $\boldsymbol{I}_{\text{morph}} = G(\boldsymbol{x}^*, \boldsymbol{c})$ with the optimal morph embedding $\boldsymbol{x}^*$ and camera parameter $\boldsymbol{c}$ corresponding to the desired view-point.

Similar to any other model the template inversion model suffers from some error in the generated face images, causing the face embedding of the generated image to not fully match the input embedding. To address this issue and to improve the generated result, we consider the GaFaR model as a face generator model and perform an iterative optimization in the input of the model so that the generated image has embedding more similar to the optimal morph embeddings. Therefore, we should solve the following optimization through an iterative process:

$$\boldsymbol{x}^*_{\text{opt}} = \underset{\boldsymbol{x}\in\mathcal{X}}{\arg\min} \, ||F(G(\boldsymbol{x}, \boldsymbol{c})) - \boldsymbol{x}^*||_2. \quad (3)$$

We consider $\boldsymbol{x}^*$ as initial value and solve this optimization with Adam optimizer with 100 iterations and learning rate of 5×10^{-3}. We can then generate face images $G(\boldsymbol{x}^*_{\text{opt}}, \boldsymbol{c})$ with any view-point. Fig. 2 illustrates the method.

3. EXPERIMENTS

We want to evalute the effectiveness of morphs generated through 3D template inversion when used for 2D morphing attacks (i.e., with a fixed frontal view point). In this context, we compare our method (**3D-Inv**) to other methods described in Section 1. We consider in particular the base inversion (**Inv**) and GAN-inversion (**GAN-Inv**) from [10], 3 StyleGAN-based methods (**SG-W** [6], **SG-W+** [5] and **MIPGAN** [7]) and one landmark-based method (**LB-Combined** [3]). All inversion-based methods are using an inverter trained against the ArcFace face recognition network [16]. The biometric loss needed for MIPGAN also uses ArcFace. Other methods are independent from any FR network. Figure 3 showcases examples of the various morphs.

We compare the effectiveness of each attack through a vulnerability analysis : 1) A source dataset is picked and pairs are selected to create morphs; 2) The morphs are enrolled into a FR system as biometric references, in order to simulate a deceptive passport registration; 3) Probes from both contributing subjects are presented to the system, trying to authenticate against the morph reference; 4) A specific operating threshold for the FR system is chosen and we evaluate the rate of successful attacks (i.e., percentage of morphs for which probes of **both** contributing subjects successfully authenticate).

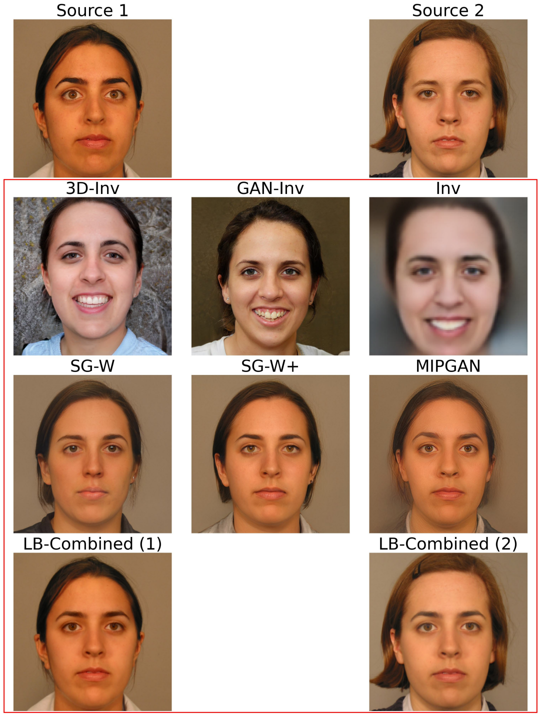

Fig. 3: Example of morphs for each attack.

In practice, we consider two source datasets. The first one is FRLL dataset [17] for which we select the same morphing pairs as in [3], using neutral & frontal views of the subjects as source images. The probes used for the vulnerability analysis are all remaining neutral images of each subject. The second one source dataset is FRGC [18]. For this one, we select both the same morphing pairs and same probes as in [7]. We run the attack against two distinct FR systems: ArcFace [16], which is already used for morph generation in the case of the inversion attacks and of MIPGAN, and ElasticFace [19]. Considering this second network enables to simulate black-box attack scenarios, i.e., the case where the attacked network is *not* available at morph generation time. For both systems, we follow the FRONTEX guideline [20] and pick an operating threshold tolerating a false-match rate of 0.1%, using protocol 2 from FRGC for the calibration. We evaluate the vulnerability of FR systems to the attacks using the Mated Morph Presentation Match Rate metric from [21], considering both the MinMax and ProdAvg variants.

Vulnerability analysis results are presented in table 1. We also evaluate (table 2) the importance of optimizing the input of the inverter by comparing the vulnerability results with and without this optimization. Finally, we evaluate in table 3 the typical runtime for generating a single morph. Keeping it low enables to generate larger morphs datasets, for example to train detection systems. Source code to reproduce our experiments is publicly available[1].

[1] Source code: https://gitlab.idiap.ch/bob/bob.paper.mlsp2025_3dmorph

Table 1: Vulnerability analysis when attacking an ArcFace system (AF) or an ElasticFace system (EF). We distinguish white-box (□) and black-box (■) attack scenarios.

		MinMax- MMPMR (%)		ProdAvg- MMPMR (%)	
FRS	Attack	FRLL	FRGC	FRLL	FRGC
AF	□ 3D-Inv	84.65	73.66	77.54	56.73
	□ Inv	**97.54**	**89.88**	**94.47**	**74.76**
	□ GAN-Inv	52.02	41.81	42.46	21.95
	□ MIPGAN	-	73.22	-	54.77
	■ SG-W	1.05	4.32	0.64	1.44
	■ SG-W+	62.63	60.10	53.71	39.97
	■ LB-Combined	**93.95**	**84.97**	**91.23**	**70.68**
EF	■ 3D-Inv	58.25	43.59	51.75	30.88
	■ Inv	91.75	75.09	86.80	58.25
	■ GAN-Inv	39.74	27.37	31.80	13.93
	■ SG-W	3.07	10.19	2.06	3.56
	■ SG-W+	72.28	67.63	62.81	48.90
	■ MIPGAN	**-**	**75.80**	**-**	**60.10**
	■ LB-Combined	**95.96**	**87.58**	**93.33**	**75.54**

Table 2: Ablation study on the impact of removing the input optimization step on the vulnerability metrics.

		MinMax- MMPMR (%)		ProdAvg- MMPMR (%)	
FRS	Attack	FRLL	FRGC	FRLL	FRGC
AF	□ 3D-Inv (no opt.)	11.40	13.65	6.78	5.31
	□ 3D-Inv	84.65	73.66	77.54	56.73
EF	■ 3D-Inv (no opt.)	10.18	12.10	6.91	5.67
	■ 3D-Inv	58.25	43.59	51.75	30.88

Table 3: End-to-end runtime of each generation algorithm to create 1 morph. Averaged over 10 morph generations.

Attack	Runtime [s]
3D-Inv	35.95 ± 0.11
3D-Inv (no opt.)	0.98 ± 0.05
Inv	0.88 ± 0.05
GAN-Inv	0.99 ± 0.05
SG-W	372.08 ± 1.46
SG-W+	373.67 ± 2.77
MIPGAN	47.43 ± 1.64
LB-Combined	0.26 ± 0.01

Visually (Fig. 3), our 3D morphs show a realism which is on par with other methods. Like other inversion-based methods, the resulting image has a typically non-uniform background and a non-neutral expression which limits its usability for passport applications. A simple cropping and blending of the face foreground into a new uniform background could solve the first issue; the second issue might depend on further developments in template inversion research to add some neutral expression constraint on the reconstructed face. Despite those limitations, we observe that the 3D morphs are of high concern in terms for FR system vulnerability (cf. table 1). In the white-box attack scenario (meaning the attacked FR system is accessible when generating the morph), the 3D-Inv method reaches MMPMR values which are competitive with MIPGAN. It is only beaten by the Inv method (which, while effective, is also the less realistic looking one) and by the LB-Combined method. This latter point follows the observation of previous literature that landmark-based morphs are still almost always more effective than deep-learning based ones. In the black-box attack scenario, the MMPMR is not as high as previous methods but is still high enough for the 3D-Inv morphs to be a concern, with notably a higher morph success rate than Inv and SG-W. In terms of generation runtime (table 3), the 3D-Inv method lies in the same order of magnitude as MIPGAN, which makes it a mid-range method, not as fast as inversion-based and landmark-based methods, but faster than all StyleGAN-based methods. We also observe that our proposed additional step of optimizing the input to the inverter in order to generate an even better morph actually matters greatly (table 2). Without this additional step, the MMPMR values decrease typically by a factor 4-10. We note that a similar idea could be applied to other inversion based methods (Inv and GAN-Inv), and hypothesize that the improvements in attack effectiveness we observe for the 3D-Inv morphing have potential to also translate to those cases.

Finally, we want to highlight that due to their 3D nature, our morphs have potential to be used in many more situations than the 2D ones, either for faking a 3D photo for a passport application using this format, like in the UAE [1], or in other unforeseen applications linked to 3D face recognition. Future work should in particular focus on comparing our 3D morphing method to the landmark-based one introduced in [15]; the different natures of the source datasets required for both methods (a point-cloud for them versus an image for us) make this comparison unfortunately non-trivial.

4. CONCLUSION

We proposed and showcased a method to produce 3D morphing attacks from 2D source images using a 3D face template inversion model. The method generates morphs of similar quality to previous methods, although its application in real-world scenarios would rely on some additional background post-processing, and expression neutralization (like for other inversion-based methods). The 3D morphs are still effective enough when attacking face recognition systems to already be of concern, especially in the white-box attack scenario where they are competitive with StyleGAN-based morphs. The generation runtime is low enough to enable generation of datasets of similar size than with StyleGAN-based morphs, which could contribute to improve the robustness of morphing attack detectors by enabling to train them on a wider variety of attacks. Moreover, the ability to generate multiple views of the morph is a large improvement over 2D methods, and is particularly interesting given the emergence of 3D passport photos like in the UAE. Instead of generating our morphs by a simple inference from the optimal morph embedding through the template inverter, we also introduce a methodology where the input to the inverter is fine-tuned by optimization to further improve the effectiveness of each individual morph; we observe that this methodology is very successful and could be applied to other inversion-based morphing attacks. A very important aspect left for future work is to compare this work to other 3D morphing attacks such as [15]; however, this comparison is non-trivial given the difference in the nature of the source datasets. They rely on 3D point-cloud data for the source faces while our method is applied to 2D datasets.

5. REFERENCES

[1] "UAE reviews features of new ID card, 3D photo included - GulfToday," https://www.gulftoday.ae/news/2021/08/05/uae-reviews-features-of-new-id-card-3d-photo-included.

[2] Matteo Ferrara, Annalisa Franco, and Davide Maltoni, "The magic passport," in *IEEE International Joint Conference on Biometrics*, Sept. 2014, pp. 1–7.

[3] T. Neubert, A. Makrushin, M. Hildebrandt, Christian Krätzer, and J. Dittmann, "Extended StirTrace benchmarking of biometric and forensic qualities of morphed face images," *IET Biometrics*, 2018.

[4] Naser Damer, Alexandra Moseguí Saladié, Andreas Braun, and Arjan 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)*, Oct. 2018, pp. 1–10.

[5] Sushma Venkatesh, Haoyu Zhang, Raghavendra Ramachandra, Kiran Raja, Naser Damer, and Christoph 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)*, Apr. 2020, pp. 1–6.

[6] Eklavya Sarkar, Pavel Korshunov, Laurent Colbois, and Sébastien Marcel, "Are GAN-based morphs threatening face recognition?," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2022, pp. 2959–2963.

[7] Haoyu Zhang, Sushma Venkatesh, Raghavendra Ramachandra, Kiran Raja, Naser Damer, and Christoph Busch, "MIPGAN—Generating Strong and High Quality Morphing Attacks Using Identity Prior Driven GAN," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 3, pp. 365–383, July 2021.

[8] Zander W Blasingame and Chen Liu, "Leveraging diffusion for strong and high quality face morphing attacks," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 6, no. 1, pp. 118–131, 2024.

[9] Una M. Kelly, Luuk Spreeuwers, and Raymond Veldhuis, "Worst-Case Morphs: A Theoretical and a Practical Approach," in *2022 International Conference of the Biometrics Special Interest Group (BIOSIG)*, Sept. 2022, pp. 1–5.

[10] Laurent Colbois, Hatef Otroshi Shahreza, and Sébastien Marcel, "Approximating optimal morphing attacks using template inversion," in *IEEE International Joint Conference on Biometric*, 2023.

[11] Hatef Otroshi Shahreza and Sébastien Marcel, "Face reconstruction from facial templates by learning latent space of a generator network," *Advances in Neural Information Processing Systems*, vol. 36, pp. 12703–12720, 2023.

[12] Hatef Otroshi Shahreza, Vedrana Krivokuća Hahn, and Sébastien Marcel, "Face reconstruction from deep facial embeddings using a convolutional neural network," in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 1211–1215.

[13] Hatef Otroshi Shahreza and Sébastien Marcel, "Template inversion attack against face recognition systems using 3d face reconstruction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023, pp. 19662–19672.

[14] Hatef Otroshi Shahreza and Sébastien Marcel, "Comprehensive vulnerability evaluation of face recognition systems to template inversion attacks via 3D face reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[15] Jag Mohan Singh and Raghavendra Ramachandra, "3d face morphing attacks: Generation, vulnerability and detection," 2024.

[16] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF*

conference on computer vision and pattern recognition, 2019, pp. 4690–4699.

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

[18] 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)*, June 2005, vol. 1, pp. 947–954 vol. 1.

[19] Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper, "ElasticFace: Elastic Margin Loss for Deep Face Recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1578–1587.

[20] FRONTEX, "Best practice technical guidelines for automated bor- der control ABC systems," 2015.

[21] Ulrich Scherhag, Andreas Nautsch, Christian Rathgeb, Marta Gomez-Barrero, Raymond N. J. Veldhuis, Luuk Spreeuwers, Maikel Schils, Davide Maltoni, Patrick Grother, Sebastien Marcel, Ralph Breithaupt, Raghavendra Ramachandra, and Christoph Busch, "Biometric Systems under Morphing Attacks: Assessment of Morphing Techniques and Vulnerability Reporting," in *2017 International Conference of the Biometrics Special Interest Group (BIOSIG)*, Sept. 2017, pp. 1–7.