Optimizing Key-Selection for Face-based One-Time Biometrics via Morphing

Dailé Osorio-Roig[1], Mahdi Ghafourian[2], Christian Rathgeb[1], Ruben Vera-Rodriguez[2], Christoph Busch[1], Julian Fierrez[2]

1 - Biometrics and Internet Security Research Group
Hochschule Darmstadt, Germany
{daile.osorio-roig,christian.rathgeb,christoph.busch}@h-da.de
2 - Biometrics and Data Pattern Analytics (BiDA) Lab
Universidad Autonoma de Madrid (UAM)
Spain
{mahdi.ghafourian,ruben.vera,julian.fierrez}@uam.es

Abstract—**Nowadays, facial recognition systems are still vulnerable to adversarial attacks. These attacks vary from simple perturbations of the input image to modifying the parameters of the recognition model to impersonate an authorised subject. So-called privacy-enhancing facial recognition systems have been mostly developed to provide protection of stored biometric reference data, i.e. templates. In the literature, privacy-enhancing facial recognition approaches have focused solely on conventional security threats at the template level, ignoring the growing concern related to adversarial attacks. Up to now, few works have provided mechanisms to protect face recognition against adversarial attacks while maintaining high security at the template level. In this paper, we propose different key selection strategies to improve the security of a competitive cancelable scheme operating at the signal level. Experimental results show that certain strategies based on signal-level key selection can lead to complete blocking of the adversarial attack based on an iterative optimization for the most secure threshold, while for the most practical threshold, the attack success chance can be decreased to approximately 5.0%.**

Index Terms—**adversarial attack, iterative optimization, face recognition, privacy protection, security, cancelable biometrics**

I. INTRODUCTION

Face recognition systems have been deployed in numerous access control applications, *e.g.* border control [1], financial transactions and ID cards [2]. However, the widespread use of these technologies has raised serious security and privacy concerns. Additionally, with the recent success of deep learning in facial recognition, potential adversarial attacks have been reported (*e.g.* [3], [4]). These attacks range from simple perturbation of the input image to advanced attacks in which model parameters are modified. [5]. According to Xu *et al.* [5], adversarial images lead to higher false match rates when security thresholds are set in a biometric system using a clean dataset (*e.g.* original face image without adversarial perturbation). In addition, when unauthorised subjects are allowed access to a restricted service or resource, they can launch adversarial attacks against the system and gain access to different applications [6], *e.g.* a genuine client's account [7].

Cancelable biometrics utilise transformations in signal or feature domain which enable a biometric comparison in the transformed (encrypted) domain [8], *i.e.* biometric templates are permanently protected. In the context of cancelable biometrics, Ghafourian *et al.* [9] proposed a scheme that aims at protecting face templates against iterative optimization-based adversarial attacks without scarifying template protection requirements [10] such as *unlinkability*, *irreversibility*, *renewability*, and *biometric performance*. More precisely, the so-called *OTB-morph* method utilises the concept of morphing attacks [11] as a transformation function for cancelable face biometrics based on time-varying keys (signal- or image-based level). The randomness employed in this transformation function (henceforth referred to as "key") is based on the random selection of the sample that contributes to a morph. Despite the fact that the method reduces the success chance of adversarial attacks produced by iterative optimization, it is still unknown to what extent the key selection (*i.e.* random selection in [9]) in OTB-morph could lead to higher security against such attacks.

Motivated by the above facts, this work investigates and proposes different key selection strategies for the OTB-morph algorithm [9]. In particular, we analyse how the probability of accepting the attacks produced by iterative optimization can be decreased by varying the selection strategy of a sample that contributes to a morph (*i.e.* key selection). While OTB-morph [9] utilises random sampling to produce a morphed face at the signal level, we exploit the properties of opposite demographic groups and dissimilarities of samples to generate morphed facial images. Said demographic properties lead to statistical assumptions already known in the literature [12]: facial recognition algorithms produce higher similarity scores and, hence, significantly more false matches for subjects sharing similar demographic attributes, *e.g.* gender and skin colour. Therefore, solutions that exploit the properties of opposing demographic groups are expected to contribute to a decrease in false matches. The findings of this work also lead to a better understanding of how signal-level cancelable facial biometrics can reduce the vulnerability of biometric systems
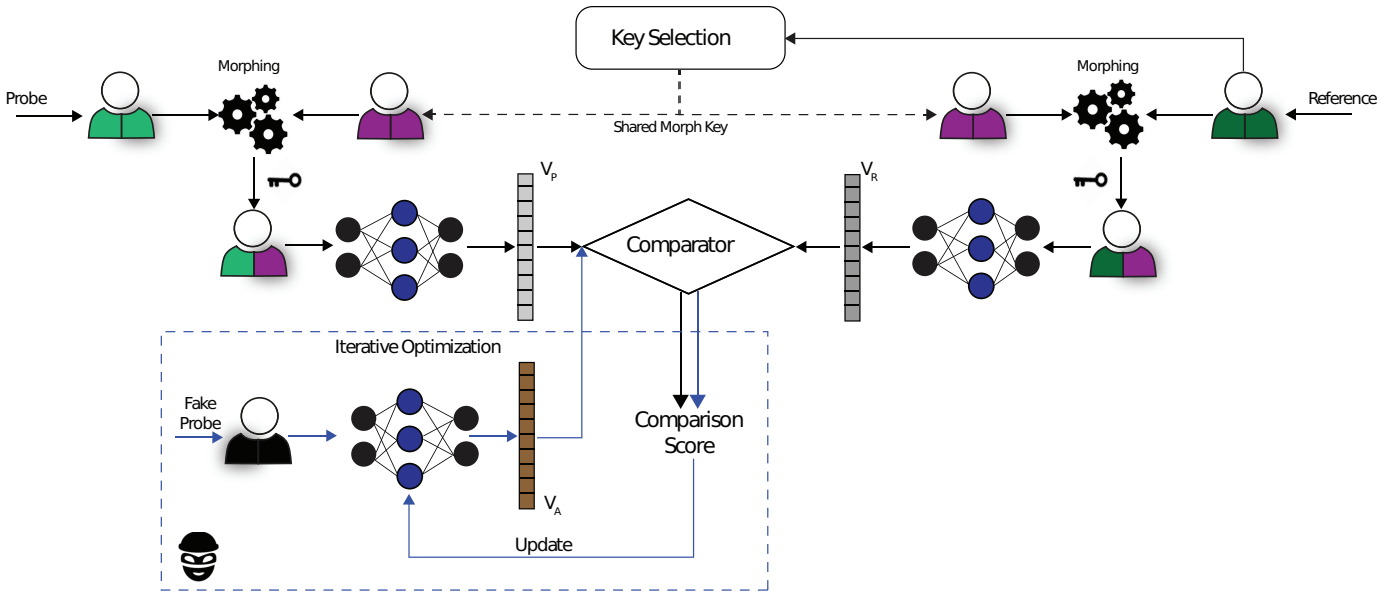
Fig. 1: Conceptual overview of key selection-based OTB-morph .

against iterative optimization-based adversarial attacks.

The remainder of this paper is organised as follows: Sect. II briefly introduces the related work. In Sect. III, a detailed explanation of OTB-morph algorithm based on different key selections is provided. Sect. IV-A presents the experimental setup and the achieved results are reported in Sect. V. A summary of the findings is finally provided in Sect. VI.

II. RELATED WORKS

This section provides a brief overview of cancelable schemes applied to biometrics. In order to improve the security and privacy of verification scenarios, the concept of cancelable biometrics was first introduced by Ratha *et al.* [13]. In particular, a cancelable face recognition system was introduced using image warping to transform biometric data in the signal domain. Until now, many other popular cancelable techniques have been developed for multiple biometric characteristics based on the application of non-invertible transformations, see [14]. Over the past years, the majority of these transformations have been improved and, most recently, cancelable transformations have been designed to work with deep neural networks (DNNs) architectures [8], [15]. Most of these approaches have been focused on the feature extraction step while preserving competitive biometric performance (*i.e.* discriminatory feature space) and high privacy protection. It is worth noting that face biometrics has recently been one of the biometric characteristics that has raised the most privacy concerns. In this context of privacy, some security gaps have been analysed on cancelable face recognition systems, *e.g.* [16], [17]. Also, several authors have addressed these gaps by introducing hybrid protection schemes. Recently, Otroshi-Shahreza *et al.* [18] investigated the hybrid protection by combining cancelable biometrics and homomorphic encryption.

Ghafourin *et al.* [9] proposed a novel time-varying cancelable scheme called OTB-Morph using the morphing concept as a cancelable transformation. The authors showed full protection of cancelable deep face templates against so-called iterative optimization-based adversarial attacks.

In summary, most of the cancelable proposals described above and existing in the literature have been analysed from different security points of view, focusing on the template level (*e.g.* [19]) and ensuring compliance with the requirements defined by the ISO/IEC 24745 standard [10]: *unlinkability*, *irreversibility*, *renewability*, and *biometric performance*. Moreover, recent research has studied adversarial attacks on state-of-the-art deep facial recognition systems [5], [20], revealing the vulnerability of facial biometric systems. To the best of the authors' knowledge and as mentioned in [9], OTB-morph has been the first work of cancelable biometric template protection scheme addressing the security threat against adversarial attacks based on iterative optimization. For comprehensive surveys on cancelable biometrics, the interested reader is referred to [8], [14], [21], [22].

III. OTB-MORPH

One-time-biometric via morphing (in short OTB-morph) is a new cancelable method to withstand iterative optimization attacks in face verification [9]. Inspired by the ultimate security of a one-time pad [23] in conventional cryptography literature, this method takes advantage of morphing as a transformation function using time-varying keys (biometrics in this case) to generate protected templates at each verification attempt (Sect. III-A). Fig. 1 shows a conceptual overview of a verification scenario protected by cancelable biometrics designed at the signal level (*i.e.* OTB-morph) that can be circumvented by an adversarial attack (*e.g.* iterative optimization-based attack). In this attack context, the non-authorised subject has

Fig. 2: Examples of morph images (2nd row) resulting from morphing a reference image (large image on the left) with each of the samples selected by the proposed key selection strategies (1st row). From left to right: Random_key, Distance_key, SFdistance_key, and SFrandom_key.

presented face images (fake probes) to the biometric system and observes the obtained comparison score and decision made by the biometric system. Finally, the iterative optimization-based attack (Sect. III-B) is injected into the system when a biometric claim is made.

A. Operation mode

An authentication attempt using OTB-morph is executed as follows: initially, a biometric claim is made; the facial image corresponding to the probe is morphed with another sample at the signal level using the OTB-morph approach, here, a key selection-based transformation function is employed signal-level morphing using the probe as the reference image; subsequently, the morphed facial image (morphed probe) is processed and a set of protected features is extracted using a deep neural network (DNN); subsequently, these features extracted from the morphed probe can be compared against features stored (a biometric reference) in the biometric system. Note that the features corresponding to the biometric reference have already been processed by the OTB-morph approach in an enrolment process. Also, it should be noted that the probe and the biometric reference share the same key selection for the morphing process. Finally, the biometric system verifies whether the claim is genuine (*i.e.* the user's identity (probe) is the one being claimed) or not (*i.e.* the user is an impostor trying to impersonate another user), and only allows access in the former case.

In our work, facial images are morphed according to different criteria to select the sample that can contribute to a morph: 1) by randomly choosing a single sample (Random_key); this type of selection has been used by OTB-Morph in the original paper [9]; 2) by selecting the most dissimilar sample (Distance_key); to that end, a dissimilarity score comparator is applied to the feature or embedding space to compute a distance (s); 3) by choosing the most dissimilar sample from the opposite demographic group (SFdistance_key); for this type of criteria, the demographic information statistics (*e.g.* gender) should be measured; 4) similar to criteria 3) but randomly selecting the sample from the opposite demographic group (SFrandom_key). Examples of resulting images for each strategy are shown in Fig. 2.

B. Iterative optimization

The idea behind this attack is to minimize a dissimilarity distance s between the victim's V_R face template (*i.e.* biometric reference) and the attacker's V_A face template (Fake Probe in the Fig. 1). Therefore, for each leaked score from an impersonation attempt, the attacker updates V_A with an adversarial perturbation p_a such that the dissimilarity score s is minimized: $\min_s |V_R - V_A|$. At each iteration, the objective function ($\min_s$) will be updated until the closest attack sample to V_R is found, *i.e.* until an impersonation attempt is successful. Note that iterative optimization-based adversarial attacks are well-known in the literature working with machine learning techniques (*e.g.* deep learning) and can be easily optimized using the gradients of DNNs, *e.g.* [24].

IV. EXPERIMENTAL SETUP

In this section, the metrics used to evaluate the different key selection strategies as well as some implementation details are summarised (Sect. IV-A). Databases and protocols employed in the assessment are also outlined (Sect. IV-B).

A. Metrics and implementation details

Similar to [9], AdaFace [25] was utilised as face feature extractor. Euclidean distance was utilised as a dissimilarity score comparator. For the morphing image process, we use the Dlib [26] implementation for landmark detection and OpenCV as the morphing tool following the same settings as in [9]. In particular, the morphing technique was applied directly to full-face images. The transformed facial images are then aligned and cropped using the open-source RetinaFace[1] software. As mentioned in Sect. III-A, these experiments take into account four different criteria for applying the morphing technique: random selection (henceforth referred to as Random_key), the most dissimilar sample (henceforth referred to as Distance_key), the most dissimilar sample from the opposite gender (*i.e.* female or male) (henceforth referred to as SFdistance_key), and random selection from the opposite gender (henceforth referred to as SFrandom_key).

The biometric performance is computed in a typical verification scenario compliant with the metrics defined in the
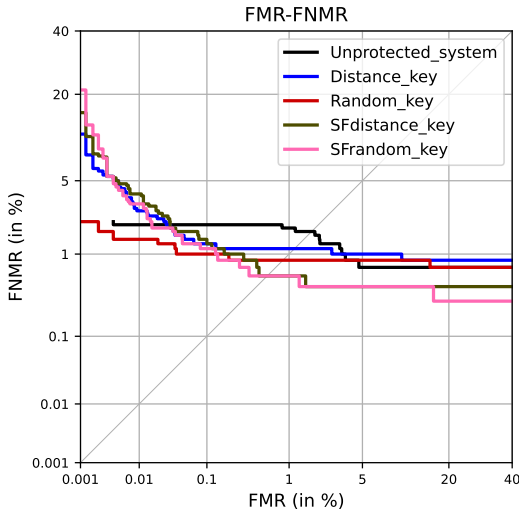
---

[1]https://github.com/serengil/retinaface

Fig. 3: Biometric performance for different key selections.

TABLE I: Error rates (in %). The best results are highlighted in bold.

System	Selection of key	EER	FMR	FNMR	Threshold	ASR
Unprotected	-	1.71	0.0010	2.00	1.1981	1.67
	-		0.0100	2.00	1.2342	6.33
	-		0.1000	2.00	1.2667	18.20
	-		1.0000	1.86	1.3074	40.73
OTB-morph	Random_key	0.86	0.0010	2.14	1.1543	0.47
			0.0100	1.42	1.1894	2.07
			0.1000	1.00	1.2292	8.93
			1.0000	0.86	1.2781	29.60
	Distance_key	1.14	0.0010	11.29	1.0751	**0.00**
			0.0100	2.71	1.1811	**0.60**
			0.1000	1.29	1.2317	**5.87**
			1.0000	1.14	1.2818	**25.33**
	SFdistance_key	0.57	0.0010	15.57	1.0370	**0.00**
			0.0100	3.86	1.1451	1.00
			0.1000	1.29	1.2043	7.60
			1.0000	0.57	1.2566	25.40
	SFrandom_key	0.57	0.0010	15.57	0.9925	**0.00**
			0.0100	3.86	1.1443	1.13
			0.1000	1.14	1.2071	8.87
			1.0000	0.57	1.2567	28.07

ISO/IEC19795-1:2021 [27] standard. The Equal Error Rate (EER), which represents the operating point at which False Match Rates (FMR) and False Non-Match Rates (FNMR) are equal, is reported. In addition, the FNMR values for several security thresholds, *i.e.* $0 \leq$ FMR ≤ 40 are depicted as Detection Error Trade-off (DET) curves. We also analysed the Attack Success Rate (ASR) which is defined by computing the number of adversarial samples that were accepted by the system at fixed security threshold.

B. Databases and protocols

Experiments are conducted on well-known face datasets such as VGGFace2 [28] and LFW [29]. The former is utilised for the biometric performance evaluation and execution of the attack. More specifically, 50 identities were selected from its test set; each identity consists of 88 samples. To evaluate biometric performance, 28 samples per identity are randomly selected, resulting in 50×14 mated comparisons and 50×91 non-mated comparisons. To conduct the attacks, the remaining 60 samples (30 references and 30 probes) are used per identity. In this case, the iterative optimization adversarial-based attack explained in Sect. III-B is performed on 30 samples from the same identity. Note that the morph key generated by key selection-based OTB-morph is changed at every face verification attempt.

LFW is used for the morphing process and the application of the different key selection criteria described in Sect. IV-A. In this context, a single sample per identity with the highest quality value estimated by the CR-FIQ framework[2] is selected, resulting in a total of 5,749 images.

V. RESULTS AND DISCUSSION

Fig. 3 benchmarks the biometric performance of different key selection strategies as well as the unprotected system

(*i.e.* baseline). Note that all proposed key selection criteria outperform the unprotected system at the most commonly used security threshold (*i.e.* FMR=0.1%). For higher security thresholds (*e.g.* FMR=0.01%), the performance yielded by all cancelable schemes is still comparable to the unprotected system.

Tab. I also shows the biometric performance, as well as the ASR values per key selection strategy and security threshold. Note that the key selection process assists in reducing the chances of attack compared to an unprotected system. In particular, for a threshold fixed at FMR=0.1%, the attack chance on protected systems is approximately seven times lower than the one achieved by the unprotected system. For stricter security thresholds, the protected scheme based on Random_key is vulnerable w.r.t. other key selections: Distance_key, SFdistance_key, and SFrandom_key report a ASR = 0% for a threshold fixed at FMR=0.001%, while a slight increase above 1.0% is observed for FMR=0.01%.

Fig. 4 reports the average dissimilarity score achieved by the attacker (*i.e.* evolutionary process) across 30 different verification attempts. Note that the dissimilarity score computed by the unprotected system gradually decreases across the iterations, thus indicating that the attacker will be accepted by the unprotected system after a few iterations or attempts for a security threshold fixed at FMR=0.1% (horizontal black line). Contrary to the trend shown by the unprotected system, no drastic changes are observed in the trend computed by the different key selection strategies. Note that such trends remain constant and above the security threshold in most iterations.

Fig. 5 shows the cumulative attack chances for the different evaluated systems across 30 iterations. It can be observed that the attack rates strongly depend on the security thresholds fixed in the system. In addition, these rates confirm the results presented in Tab. I: the attack chances are reduced to 0% for most of the key selection strategies at FMR=0.001% with the exception of Random_key (green line in Fig. 5a). More im-
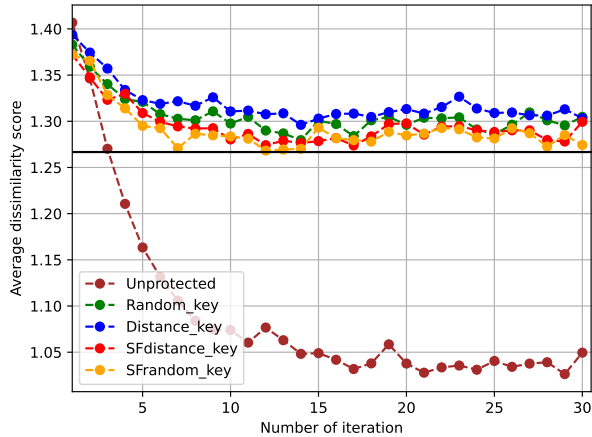
Fig. 4: Evolution of the average comparison score achieved by the attacker. Horizontal black line visualizes the security threshold fixed at FMR=0.1% in the baseline (*i.e.* unprotected system).

portantly, the constant zero behaviour (*i.e.* complete blocking of the attack) for all attempts can be observed, in contrast to Random_key. For FMR=0.01% (Fig. 5b), the attack chance appears to be constant from a certain number of iterations for some key selections (*e.g.* the attack is constant from iteration 15 for Distance_key (blue line), SFdistance_key (red line), and SFrandom_key (yellow line)). For FMR=0.1% (Fig. 5c), Distance_key (blue line) appears to be more promising, while for the more relaxed thresholds (*i.e.* for FMR=1.0% (Fig. 5d) and FMR=FNMR (Fig. 5e), the key selection based on opposite demographic information (red and yellow lines) is more challenging for the attacker. In summary, for the recommended security threshold of FMR=0.1%, the most dissimilar image (*i.e.* Distance_key) is the best choice to be used as a key in the morphing process.

VI. SUMMARY

This work has shown that cancelable biometrics working at the signal level can be resistant to adversarial attacks. More specifically, new defence mechanisms in key selection strategies working on morphing techniques were shown to drastically reduce the chances of impostors (*e.g.* impersonation attempts) produced by the iterative optimization-based attack. An empirical evaluation (OTB-morph in this case) showed that the randomness of signal-level cancelable schemes does not usually circumvent such attacks at their optimum. Here, the knowledge of demographic information and score distances reduced the chances of attack success down to zero percent for the highest security levels in a protected face recognition system. Future work will be focused on the impact of key selection-based OTB-morph on the variation of the dissimilarity function and face embedding extractors.
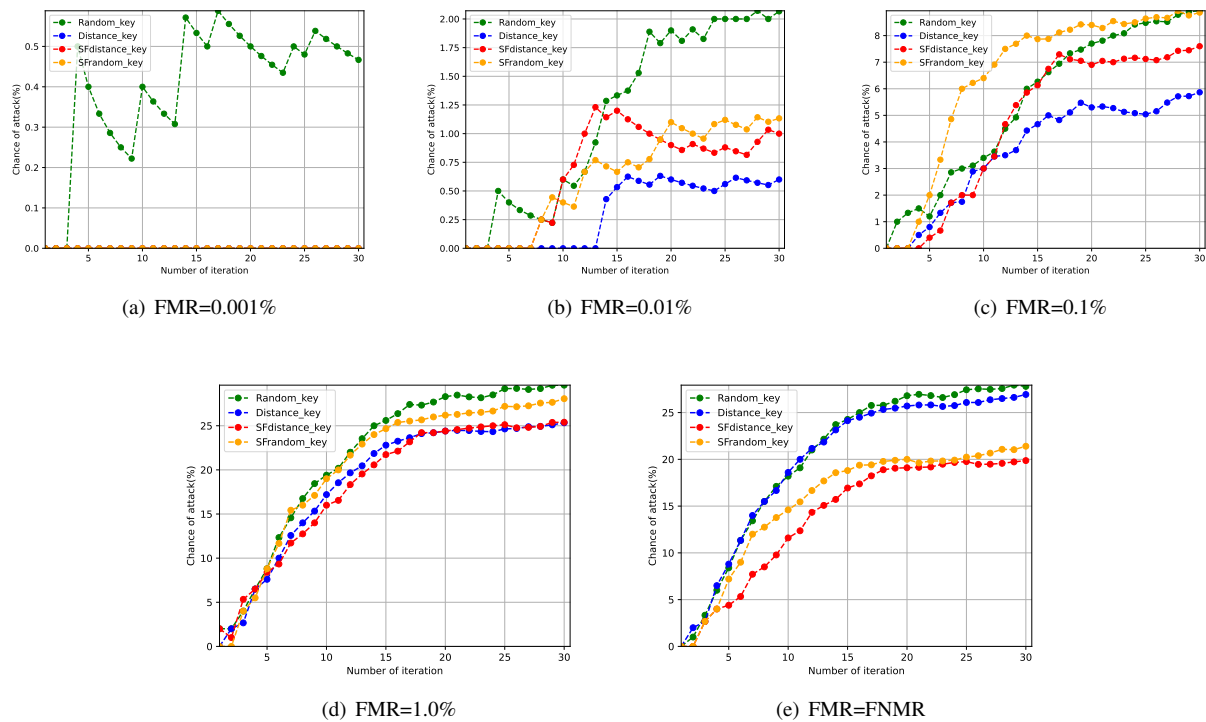
REFERENCES

[1] European Commission, "Smart borders," https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/smart-borders-en, 2018, last accessed: September 25, 2023.

[2] M. William, "National id cards launched in drc, remain uncollected in ghana," https //www.biometricupdate.com/202307/national-id-cards-launched-in-drc-remain-uncollected-in-ghana, Jul.2023, last accessed: September 25, 2023.

[3] A. Agarwal, R. Singh, M. Vatsa, and N. Ratha, "Are image-agnostic universal adversarial perturbations for face recognition difficult to detect?" in *2018 IEEE 9th Intl. Conf. on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2018, pp. 1–7.

[4] A. Dabouei, S. Soleymani, J. Dawson, and N. Nasrabadi, "Fast geometrically-perturbed adversarial faces," in *2019 IEEE Winter Conf. on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 1979–1988.

[5] Y. Xu, K. Raja, R. Ramachandra, and C. Busch, "Adversarial attacks on face recognition systems," in *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*. Springer International Publishing Cham, 2022, pp. 139–161.

[6] B. Biggio, P. Russu, L. Didaci, F. Roli *et al.*, "Adversarial biometric recognition: A review on biometric system security from the adversarial machine-learning perspective," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 31–41, 2015.

[7] B. Biggio, G. Fumera, and F. Roli, "Security evaluation of pattern classifiers under attack," *IEEE transactions on knowledge and data engineering*, vol. 26, no. 4, pp. 984–996, 2013.

[8] C. Rathgeb, J. Kolberg, A. Uhl, and C. Busch, "Deep learning in the field of biometric template protection: An overview," *arXiv preprint arXiv:2303.02715*, 2023.

[9] M. Ghafourian, J. Fierrez, R. Vera-Rodriguez, A. Morales, and I. Serna, "Otb-morph: One-time biometrics via morphing," *Machine Intelligence Research*, pp. 1–17, 2023.

[10] *ISO/IEC 24745:2022(E) Information technologyy, cybersecurity and privacy protection – Biometric information protection*, International Organization for Standardization International Standard, Feb. 2022.

[11] 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. Raghavendra, and C. Busch, "Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting," in *Intl. Conf. of the Biometrics Special Interest Group BIOSIG 2017*, 2017, pp. 1–7.

[12] J.-J. Howard, Y.-B. Sirotin, and A.-R. Vemury, "The effect of broad and specific demographic homogeneity on the imposter distributions and false match rates in face recognition algorithm performance," in *2019 IEEE 10th Intl. Conf. on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2019, pp. 1–8.

[13] N.-K. Ratha, J.-H. Connell, and R.-M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.

[14] V.-M. Patel, R.-N. Ratha, and R. Chellappa, "Cancelable biometrics: A review," *IEEE signal processing magazine*, vol. 32, no. 5, pp. 54–65, 2015.

[15] H.-O. Shahreza, V.-K. Hahn, and S. Marcel, "On the recognition performance of biohashing on state-of-the-art face recognition models," in *2021 IEEE Intl. Workshop on Information Forensics and Security (WIFS)*. IEEE, 2021, pp. 1–6.

(a) FMR=0.001%

(b) FMR=0.01%

(c) FMR=0.1%

(d) FMR=1.0%

(e) FMR=FNMR

Fig. 5: Cumulative attack chances across different numbers of iterations. Note that FMR=FNMR is defined in the threshold fixed at the EER.

[16] H. Wang, X. Dong, Z. Jin, A.-B.-J. Teoh, and M. Tistarelli, "Interpretable security analysis of cancellable biometrics using constrained-optimized similarity-based attack," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 70–77.

[17] L. Ghammam, K. Karabina, P. Lacharme, and K. Thiry-Atighehchi, "A cryptanalysis of two cancelable biometric schemes based on index-of-max hashing," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2869–2880, 2020.

[18] H. Otroshi-Shahreza, C. Rathgeb, D. Osorio-Roig, V. Krivokuća, S. Marcel, and C. Busch, "Hybrid protection of biometric templates by combining homomorphic encryption and cancelable biometrics," in *Proc. of the 2022 Intl. Joint Conf. on Biometrics (IJCB)*. IEEE, October 2022.

[19] X. Dong, Z. Jin, A. Teoh, M. Tistarelli, and K. Wong, "On the security risk of cancelable biometrics," *arXiv preprint arXiv:1910.07770*, 2019.

[20] F. Vakhshiteh, A. Nickabadi, and R. Ramachandra, "Adversarial attacks against face recognition: A comprehensive study," *IEEE Access*, vol. 9, pp. 92 735–92 756, 2021.

[21] Manisha and N. Kumar, "Cancelable biometrics: A comprehensive survey," *Artificial Intelligence Review*, vol. 53, pp. 3403–3446, 2020.

[22] C. Rathgeb and A. Uhl, "A survey on biometric cryptosystems and cancelable biometrics," *EURASIP Journal on Information Security*, vol. 3, March 2011.

[23] F. Rubin, "One-time pad cryptography," *Cryptologia*, vol. 20, no. 4, pp. 359–364, 1996.

[24] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," *Advances in neural information processing systems*, vol. 32, 2019.

[25] M. Kim, A. Jain, and X. Liu, "Adaface: Quality adaptive margin for face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18 750–18 759.

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

[27] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 19795-1:2021. Information Technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework*, International Organization for Standardization, June 2021.

[28] Q. Cao, L. Shen, W. Xie, O. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, 2018, pp. 67–74.

[29] G. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Faces in the wild: a database for studying face recognition in unconstrained environments," *Technical Report*, pp. 07–49, 2007.