On the Recognition Performance of BioHashing on state-of-the-art Face Recognition models

Hatef Otroshi Shahreza*†, Vedrana Krivokuća Hahn*, and Sébastien Marcel*‡
*Biometrics Security and Privacy Group, Idiap Research Institute, Switzerland
†School of Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
‡School of Criminal Justice, University of Lausanne (UNIL), Switzerland

***Abstract*—Face recognition has become a popular authentication tool in recent years. Modern state-of-the-art (SOTA) face recognition methods rely on deep neural networks, which extract discriminative features from face images. Although these methods have high recognition performance, the extracted features contain privacy-sensitive information. Hence, the users' privacy would be jeopardized if the features stored in the face recognition system were compromised. Accordingly, protecting the extracted face features (templates) is an essential task in face recognition systems. In this paper, we use BioHashing for face template protection and aim to establish the *minimum* BioHash length that would be required in order to maintain the recognition accuracy achieved by the corresponding unprotected system. We consider two hypotheses and experimentally show that the performance depends on the value of the BioHash length (as opposed to the ratio of the BioHash length to the dimension of the original features). To eliminate bias in our experiments, we use several SOTA face recognition models with different network structures, loss functions, and training datasets, and we evaluate these models on two different datasets (LFW and MOBIO). We provide an open-source implementation of all the experiments presented in this paper so that other researchers can verify our findings and build upon our work.**

***Index Terms*—BioHashing, Biometrics, deep features, Face recognition, Template protection**

I. Introduction

Biometric recognition systems are widely used and growing in different applications. Among biometric recognition systems, face recognition systems have become the most popular and prevalent. With the recent advances in deep learning, the state-of-the-art (SOTA) face recognition methods are mainly based on deep neural networks (DNNs). In such systems, a convolutional neural network (CNN) is used to extract features, called "embeddings", from face images. These deep features are stored in the system's database during the enrollment stage and are later used for recognition. Therefore, the features stored in the system's database contain critical information about the users' identities [1]. Since biometric traits, like face, are unique and cannot be changed, the users' privacy would be jeopardized if their biometric features were compromised [2]. To tackle privacy issues in biometric recognition systems, several biometric template protection (BTP) methods have been proposed in the literature [2], [3]. Generally, a biometric template protection scheme should have four main properties [4], [5]:

- *Recognition Performance*: Template protection should not result in recognition accuracy degradation, meaning that the protected templates should have accurate recognition.
- *Irreversibility*: It should be computationally impossible to reconstruct the original biometric data from the protected templates.
- *Cancelability*: We should be able to revoke the enrolled protected templates if they are compromised and replace them with new protected templates.
- *Unlinkability*: Considering the cancelability property, there should be no link between different protected templates from the same original biometric feature.

One of the most popular categories of BTP methods is called *cancelable* template protection methods, in which a transformation function, which is dependent on a *key* [6], is often used. Hence, by changing the key, a new protected template can be generated for the same biometric feature. BioHashing [7] is one of the most widely studied and well-known cancelable template protection methods. In particular, BioHashing has been shown to be applied to various biometric characteristics (e.g. finger vein [8]–[10], fingerprints [7], iris [11], face [12], palm prints [13]).

In this paper, we use the BioHashing scheme to generate protected templates from SOTA face recognition models. We evaluate the recognition performance of BioHashing in two scenarios: the *normal* scenario (which is the expected scenario in practice) and the *stolen* scenario (which is the case when the user's BioHashing key is stolen). We use two datasets, Labeled Faced in the Wild (LFW) [14] and MOBIO [15], to evaluate several SOTA face recognition models, which vary in their network structure, loss function, and training dataset.

When designing a BioHash-protected system, an important decision is the appropriate *length* (i.e., number of bits) of the protected templates. In this paper, we aim to establish the *minimum* BioHash length that would be required in order for our BioHash-protected face recognition systems to maintain the recognition accuracy achieved by their corresponding baselines (i.e., unprotected systems). To this end, we define and evaluate two hypotheses on how the BioHash length affects the performance of our BioHash-protected face recognition systems. The first hypothesis assumes that the performance of a BioHash-protected system depends on the *ratio* of the BioHash length to the dimensionality of the original feature vector. The second hypothesis assumes that the performance

This research is based upon work supported by the H2020 TReSPAsS-ETN Marie Skłodowska-Curie early training network (grant agreement 860813).

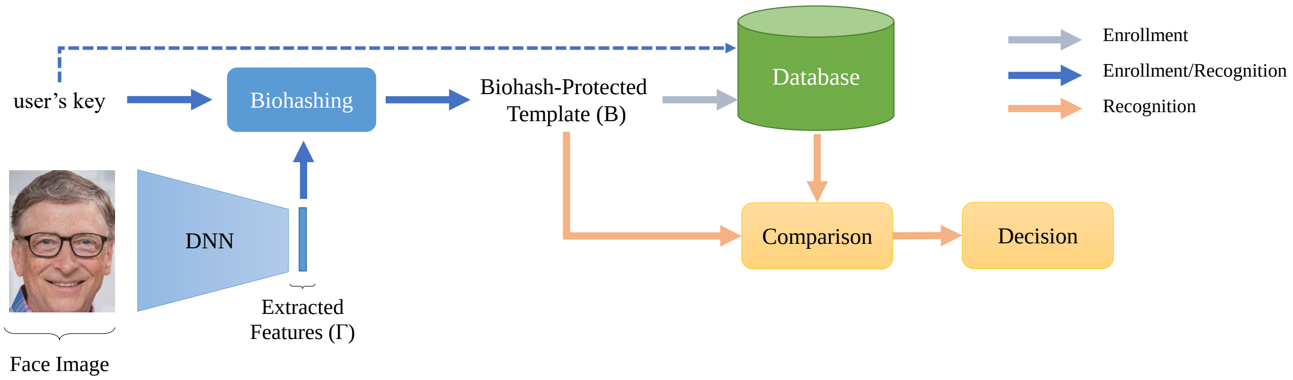


Fig. 1: Block diagram of a Biohash-protected Face Recognition system (ISO/IEC 30136)

depends on the value of the BioHash length (as opposed to the ratio), thereby we expect to see similar behavior over different models when they are protected with the same BioHash length. Our experiments on different face recognition models show that, in general, the first hypothesis is false and the second hypothesis is true. To the best of our knowledge, these findings concerning the effect of the BioHash length on the recognition accuracy of deep-learning-based face recognition models, represent a new contribution to the field of biometric template protection.

The rest of this paper is organized as follows. First, we describe the BioHash-protected system and our hypotheses in section II. Next, we present our experiments and discuss the results in section III. Finally, the paper is concluded in section IV.

II. Methodology

As mentioned in section I, in this paper we evaluate the recognition performance of BioHashing [7] on SOTA face recognition models. To this end, we considered a face recognition system following ISO/IEC 30136 [5] as illustrated in figure 1, and we used BioHashing (as described in section II-A) to generate protected templates from the features extracted by SOTA face recognition DNN models.

In section II-B, we define two hypotheses on how BioHash length can affect the performance of BioHash-protected system. According to our hypotheses, we consider different experiments and evaluate the performance of different BioHash-protected face recognition systems to verify our hypotheses. We should note that we evaluate the performance of BioHash-protected versions of the face recognition methods, and we do not evaluate the irreversibility, cancelability, and unlinkablity characteristics of BioHashing since the evaluation of these characteristics have been studied in the literature [7], [16]–[20].

A. BioHashing algorithm

Let's consider Γ, indicating an unprotected biometric template (i.e., embeddings) extracted by a face recognition model. The BioHash-protected template, B, can be generated by algorithm 1 using the user's key, k, and their features, Γ [7].

Algorithm 1 BioHashing algorithm

```
Inputs:
    $\Gamma$ : unprotected biometric template (i.e., embeddings)
    $L_e$ : length of the unprotected template ($\Gamma$)
    $L_b$ : length of the BioHash-protected template
    $k$ : user's seed
Output: $B = \{b_i | i = 1, 2, ..., L_b\}$ binary BioHash protected template
Procedure:
Generate a set of pseudo-random vectors, $\{r_i \in \mathbb{R}^{L_e} | i = 1, 2, ..., L_b\}$, based on the user's seed, $k$.
Apply the Gram-Schmidt process to transform the generated pseudo-random vectors $\{r_i \in \mathbb{R}^{L_e} | i = 1, ..., L_b\}$ into an orthonormal set of vectors $\{r_{\perp i} \in | i = 1, ..., L_b\}$
Compute $\{\langle \Gamma, r_{\perp i} \rangle \in \mathbb{R} | i = 1, ..., L_b\}$ where $\langle ., . \rangle$ indicates inner product operation.
Compute $L_b$ bits BioHash $\{b_i | i = 1, 2, ..., L_b\}$ from
    $b_i = \begin{cases} 0 & \text{if } \langle \Gamma, r_{\perp i} \rangle \leq \tau \\ 1 & \text{if } \langle \Gamma, r_{\perp i} \rangle > \tau \end{cases}, i = 1, ..., L_b,$
    where $\tau$ is a preset threshold.
End Procedure
```

During the enrollment stage, the BioHashed templates, B, and the user's key should be stored in the system database (ideally separately). During the recognition stage, Hamming distance is used to find the score between each pair of probe and reference BioHashed template. In the subsequent experiments, we consider the BioHash-protected face recognition systems operating in verification mode only.

B. Hypotheses

Let's consider a BioHash-protected face recognition system with original dimension L_e, BioHash length L_b and the ratio $\alpha = L_b/L_e$. We define two hypotheses on the performance of BioHash-protected systems:

Hypothesis 1: *The performance of BioHash-protected templates depends on α.*

According to this hypothesis, we expect to have similar trend over different face recognition models when we have similar values for α. To evaluate this hypothesis, we evaluate different models with different L_e and see if we have the same behavior with similar values for α. In such experiments, we use the value of BioHash length $L_b = \alpha \times L_e$, so that we have desired α with respect to L_e.

Hypothesis 2: *The performance of BioHash-protected templates depends on* L_b.

Based on the second hypothesis, we expect to see similar behavior over different models when they are protected with the same L_b.

According to the aforementioned hypotheses, the performance of BioHash-protected system either depends on α or exact value of L_b. In the next section, we evaluate various face recognition models to verify which hypothesis is correct in practice.

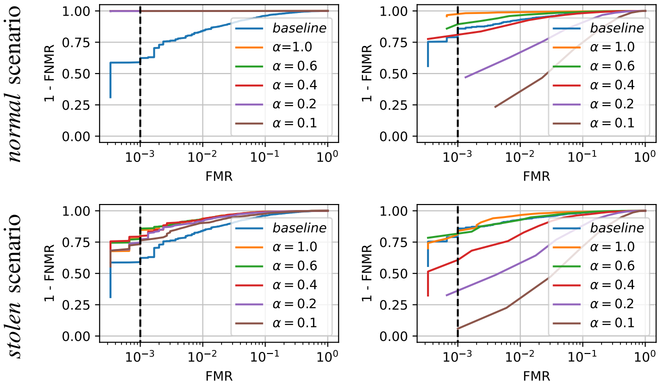
III. Experiments

In this section, we describe our experiments and evaluate the performance of different BioHash-protected face recognition systems to verify our hypotheses described in section II-B. First, in section III-A we describe our experimental setup and different baselines used in our experiments. Next, we evaluate *hypothesis 1* and *hypothesis 2* in sections III-B and III-C, respectively. Finally, we discuss our experiments and conclude our findings in section III-D.

A. Experimental Setup and Baselines

As stated in section I, in our experiments we used the Labeled Faced in the Wild (LFW) [14] and MOBIO [15] databases to evaluate the recognition performance of BioHashing on SOTA face recognition models. The LFW database includes 13,233 images of 5,749 people, where 1,680 people have two or more images. We used the *View 2* protocol[1] to evaluate the models. The MOBIO dataset is a bimodal dataset including face and audio data taken with mobile devices from 152 people. We used the *development* subset of *mobio-all* protocol[2] in our experiments.

We also used several SOTA face recognition models[3] including VGG-Oxford [21], AFFFE [22], ArcFace-InsightFace[4] [23], FaceNet[5] [24], IncResNetV1-MSCeleb1M-CenterLoss[6] [25], IncResNetV2-MSCeleb1M-CenterLoss[7] [25] ResNet50-VGG2-ArcFace[8] [26], and MobileNetV2-MSCeleb1M-ArcFace[9].



(a)

(b)

Fig. 2: ROC curves of the unprotected (i.e., baseline) and BioHash-protected versions of features extracted by a) VGG-Oxford ($L_e = 4096$) and b) FaceNet ($L_e = 128$) models with different values of $\alpha = L_b/L_e$ on LFW dataset in the *normal* (first row) and *stolen* (second row) scenarios

As mentioned in Section I, in our experiments we considered two scenarios: the *normal* scenario and the *stolen token* scenario. In the *normal* scenario, which is the expected scenario for most cases, each user's key is assumed to be secret. However, in the *stolen token* scenario (or briefly *stolen* scenario), we assumed that the impostor has access to the genuine user's secret key and uses this key with the impostor's own face features. While such a scenario is expected to happen rarely in practice, the security of the BioHash-protected face recognition system relies on the secrecy of each user's key. To implement the *stolen* scenario, in the verification stage we used the same key as the genuine's key for other users in the database and calculated the BioHash codes.

For our experiments, we used the Bob[10] toolbox [27], [28] and the open-source implementation of the BioHashing algorithm in Bob[11] [8], [9]. The source code and the trained models from our experiments are publicly available to help reproduce our results[12].

B. Evaluation of Hypothesis 1

In order to evaluate our first hypothesis, we plotted the Receiver Operating Characteristic (ROC) curve for different values of α, for each of our BioHash-protected face recognition systems. Figure 2 illustrates the corresponding plots for the VGG-Oxford and FaceNet systems. As this figure

[1] The implementation of *View 2* protocol for the LFW dataset is available at https://gitlab.idiap.ch/bob/bob.db.lfw

[2] The implementation of the *mobio-all* protocol for the MOBIO dataset is available at https://gitlab.idiap.ch/bob/bob.db.mobio

[3] The implementation of each face recognition model is available at https://gitlab.idiap.ch/bob/bob.bio.face

[4] ArcFace model with LResNet100 backbone from InsightFace: https://github.com/deepinsight/insightface

[5] FaceNet model 20170512_110547 from David Sanderberg: https://github.com/davidsandberg/facenet

[6] CNN with Inception-ResNet-v1 backbone trained on the MS-Celeb-1M dataset with the cross-entropy loss and center loss.

[7] CNN with Inception-ResNet-v2 backbone trained on the MS-Celeb-1M dataset with the cross-entropy loss and center loss.

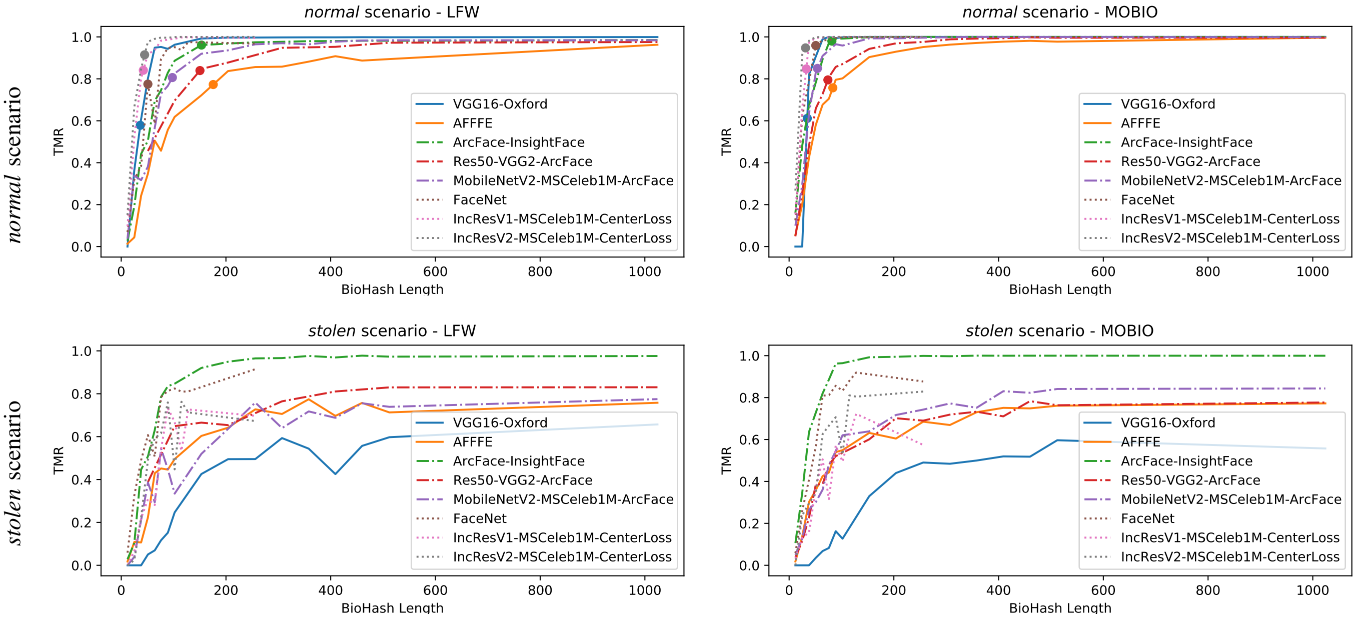
[8] CNN with Resnet-50 backbone trained on the VGG2Face dataset with ArcFace loss.

[9] CNN with MobileNet-V2 backbone trained on the MS-Celeb-1M dataset with ArcFace loss.

[10] https://www.idiap.ch/software/bob/

[11] https://gitlab.idiap.ch/bob/bob.paper.tbiom2021_protect_vascular_dnn_biohash

[12] Source code: https://gitlab.idiap.ch/bob/bob.paper.wifs2021_biohashing_sota_face



(a) LFW dataset

(b) MOBIO dataset

Fig. 3: Diagram of TMR at FMR $= 10^{-3}$ vs BioHash length (L_b) for different BioHash-protected face recognition models in both *normal* (first row) and *stolen* (second row) scenarios on LFW (a) and MOBIO (b) databases. For the normal scenario, the marked points indicate where the TMR of BioHash-protected system reaches $0.98 \times$ TMR of the corresponding baseline. The dotted lines represent models with $L_e = 128$, the dash-dotted lines show models with $L_e = 512$, and models with $L_e > 512$ are shown with solid lines (i.e., VGG-Oxford [$L_e = 4096$] and AFFFE [$L_e = 1000$]).

shows, the performance of BioHash-protected systems does not necessarily depend on the value of α across different face recognition models, and therefore *hypothesis 1* is not correct in general. For example, we observe that while $\alpha = 0.1$ has significantly degraded the performance of BioHash-protected FaceNet system ($L_e = 128$), the same α achieves competitive performance with the BioHash-protected VGG-Oxford systems ($L_e = 4096$) with large values of α and its corresponding baseline. We may thus conclude that hypothesis 1 is rejected in general.[13].

C. Evaluation of Hypothesis 2

To verify hypothesis 2, we evaluated the performance of our BioHash-protected systems (using different face recognition models) for different values of the BioHash length. We used a match threshold at a False Match Rate (FMR) of 10^{-3}, and we calculated the True Match Rate (TMR)[14] for each system. Figure 3 illustrates the resulting TMR vs L_b for the different face recognition models on which our BioHash-protected systems are based, for the LFW and MOBIO databases. For the normal scenario plots, we also marked the points where the TMR of each BioHash-protected system reaches $0.98 \times$ TMR of its corresponding baseline[15]. These points are summarized in Table I, in terms of their L_b (denoted as $L_b{}^*$) and α (denoted as α^*). For comparison, Table I also presents the number of features in the unprotected face embeddings (L_e) corresponding to each baseline face recognition system, as well as the baseline performance in terms of TMR on the LFW and MOBIO datasets. As in figure 3, , the threshold for the TMR results in table I for each system was selected individually at FMR $= 10^{-3}$. As figure 3 shows, the marked points are close and their BioHash lengths (as reported in table I) are also in the same range. These results can be explained by *hypothesis 2*.

D. Discussions

In section III-B, we used the ROC curves and compared BioHash-protected systems with different values of α and concluded that *hypothesis 1* can not be true in general. Table I also shows that the values of α for the BioHash-protected systems which reach $0.98 \times$ TMR of their corresponding baselines have a large range from 0.01 to 0.40. Hence, this table also support our conclusion on rejecting hypothesis 1.

Table I also shows that the values of $L_b{}^*$ are in range 31 to 175. This shows that even for baselines with large embeddings (e.g., 4096 for VGG-Oxford, or 1000 for AFFFE),

[13] The corresponding plots for other models are also available in the software package of the paper.

[14] TMR $= 1 -$ FNMR (where FNMR is False Non-Match Rate).

[15] To compensate for the models with very well-performing baselines, we consider achieving $0.98 \times$ TMR baseline as maintaining the baseline performance.

TABLE I: Comparison of SOTA Face Recognition models and their corresponding L_b and α where TMR of BioHash-protected system reaches $0.98\times$TMR$_{\text{baseline}}$ on LFW and MOBIO datasets. The thresholds for the baselines and BioHash-protected systems are selected individually at a False Match Rate (FMR) of 10^{-3}

Model	$\mathbf{L_e}$	**LFW**			**MOBIO**		
		TMR$_{\text{baseline}}$	$\mathbf{L_b}^*$	α^*	TMR$_{\text{baseline}}$	$\mathbf{L_b}^*$	α^*
VGG-Oxford	4096	0.59	36	0.01	0.62	34	0.01
AFFFE	1000	0.79	175	0.34	0.77	83	0.16
ArcFace-InsightFace	512	0.98	153	0.30	1.00	81	0.16
ResNet50-VGG2-ArcFace	512	0.81	74	0.14	0.86	150	0.29
MobileNetV2-MSCeleb1M-ArcFace	512	0.87	54	0.11	0.82	97	0.19
FaceNet	128	0.98	50	0.40	0.79	50	0.40
IncResNetV1-MSCeleb1M-CenterLoss	128	0.86	32	0.26	0.86	42	0.33
IncResNetV2-MSCeleb1M-CenterLoss	128	0.97	31	0.24	0.93	44	0.35

${L_b}^*$ and α^* correspond to the BioHash-protected system which its TMR reaches $0.98 \times$ TMR$_{\text{baseline}}$

choosing the BioHash length around 175 helps to maintain the recognition performance of the corresponding baseline.

We should also note that as depicted in figure 3, in the case of the *normal* scenario, BioHash-protected templates with large values of BioHash length can outperform unprotected templates (baseline). However, in the *stolen* scenario, BioHashing degrades the performance. This can be interpreted with respect to the secret or the disclosed key in these scenarios. In the *normal* scenario, using a secret key for each user helps to increase the distance between the protected templates from different users, thereby better distinguishing between templates of users. This is not the case when the key is revealed as in the *stolen* scenario.

Last but not least, we should note that in our experiments, we used different face recognition methods with numbers of extracted features, network structures, loss functions, and training datasets. Besides, we evaluated the models on two different challenging datasets (LFW and MOBIO). Therefore, there are no inherent assumptions or inherent biases regarding network structures, loss functions, training data, or test data in our experimental results.

IV. Conclusion

In this paper, we used BioHashing to protect SOTA face recognition methods and evaluated their performance in *normal* and *stolen* scenarios. We aimed to find the *minimum* value of BioHash length that would maintain the recognition accuracy of the corresponding unprotected face recognition system. To this end, we considered two hypotheses on how BioHash length affects the performance of BioHash-protected system. Our experiments show that in general, the performance depends only on the BioHash length (as oppose to the ratio of BioHash length to the dimensionality of the original feature vector). To prevent bias in our experiments, we evaluated the models on two different datasets (LFW and MOBIO) and used various face recognition methods with different network structure, loss functions, and training datasets.

V. Acknowledgment

The authors would like to acknowledge Tiago de Freitas Pereira (Idiap Research Institute, Switzerland) and Manuel Günther (University of Zurich, Switzerland) for their help and supports in the implementation of baselines of face recognition models.

References

[1] G. Mai, K. Cao, P. C. Yuen, and A. K. Jain, "On the reconstruction of face images from deep face templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 5, pp. 1188–1202, 2018.

[2] K. Nandakumar and A. K. Jain, "Biometric template protection: Bridging the performance gap between theory and practice," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 88–100, 2015.

[3] A. Sarkar and B. K. Singh, "A review on performance, security and various biometric template protection schemes for biometric authentication systems," *Multimedia Tools and Applications*, pp. 1–56, 2020.

[4] G. Mai, K. Cao, X. Lan, and P. C. Yuen, "Secureface: Face template protection," *IEEE Transactions on Information Forensics and Security*, 2020.

[5] *ISO/IEC 30136:2018(E) Information technology – Security techniques – Performance testing of biometric template protection schemes*, International Organization for Standardization International Standard, Jun. 2018.

[6] M. Sandhya and M. V. Prasad, "Biometric template protection: A systematic literature review of approaches and modalities," in *Biometric Security and Privacy*. Springer, 2017, pp. 323–370.

[7] A. T. B. Jin, D. N. C. Ling, and A. Goh, "Biohashing: two factor authentication featuring fingerprint data and tokenised random number," *Pattern Recognition*, vol. 37, no. 11, pp. 2245–2255, 2004.

[8] H. O. Shahreza and S. Marcel, "Towards protecting and enhancing vascular biometric recognition methods via biohashing and deep neural networks," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 3, pp. 394–404, 2021.

[9] V. Krivokuća and S. Marcel, "On the recognition performance of biohash-protected finger vein templates," in *Handbook of Vascular Biometrics*. Springer, Cham, 2020, pp. 465–480.

[10] H. O. Shahreza and S. Marcel, "Deep auto-encoding and biohashing for secure finger vein recognition," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 2585–2589.

[11] C. S. Chin, A. T. B. Jin, and D. N. C. Ling, "High security iris verification system based on random secret integration," *Computer Vision and Image Understanding*, vol. 102, no. 2, pp. 169–177, 2006.

[12] A. Goh and D. C. Ngo, "Computation of cryptographic keys from face biometrics," in *IFIP International Conference on Communications and Multimedia Security*. Springer, 2003, pp. 1–13.

[13] T. Connie, A. Teoh, M. Goh, and D. Ngo, "Palmhashing: a novel approach for cancelable biometrics," *Information Processing Letters*, vol. 93, no. 1, pp. 1–5, 2005.

[14] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, October 2007.

[15] C. McCool, R. Wallace, M. McLaren, L. El Shafey, and S. Marcel, "Session variability modelling for face authentication," *IET Biometrics*, vol. 2, no. 3, pp. 117–129, Sep. 2013.

[16] K. H. Cheung, A. W.-K. Kong, J. You, D. Zhang *et al.*, "An analysis on invertibility of cancelable biometrics based on biohashing." in *CISST*, vol. 2005. Citeseer, 2005, pp. 40–45.

[17] R. Belguechi, E. Cherrier, and C. Rosenberger, "How to evaluate transformation based cancelable biometric systems?" in *NIST International Biometric Performance Testing Conference (IBPC)*, 2012.

[18] A. Nagar, K. Nandakumar, and A. K. Jain, "Biometric template transformation: a security analysis," in *Media Forensics and Security II*, vol. 7541. International Society for Optics and Photonics, 2010, p. 75410O.

[19] Y. Lee, Y. Chung, and K. Moon, "Inverse operation and preimage attack on biohashing," in *IEEE Workshop on Computational Intelligence in Biometrics: Theory, Algorithms, and Applications*. IEEE, 2009, pp. 92–97.

[20] X. Dong, Z. Jin, and A. T. B. Jin, "A genetic algorithm enabled similarity-based attack on cancellable biometrics," in *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2019, pp. 1–8.

[21] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," 2015.

[22] C. Li, M. Gunther, and T. E. Boult, "Eclipse: Ensembles of centroids leveraging iteratively processed spatial eclipse clustering," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, pp. 131–140.

[23] J. Deng, J. Guo, X. Niannan, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

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

[25] T. de Freitas Pereira, A. Anjos, and S. Marcel, "Heterogeneous face recognition using domain specific units," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 7, pp. 1803–1816, 2018.

[26] T. de Freitas Pereira and S. Marcel, "Fairness in biometrics: a figure of merit to assess biometric verification systems," *arXiv e-prints*, pp. arXiv–2011, 2020. [Online]. Available: https://arxiv.org/abs/2011.02395v2

[27] A. Anjos, L. E. Shafey, R. Wallace, M. Günther, C. McCool, and S. Marcel, "Bob: a free signal processing and machine learning toolbox for researchers," in *Proceedings of the 20th ACM Conference on Multimedia Systems (ACMMM)*, Oct. 2012.

[28] A. Anjos, M. Günther, T. de Freitas Pereira, P. Korshunov, A. Mohammadi, and S. Marcel, "Continuously reproducing toolchains in pattern recognition and machine learning experiments," in *Proceedings of the International Conference on Machine Learning (ICML)*, Aug. 2017.