Stable Hash Generation for Efficient Privacy-Preserving Face Identification

Dailé Osorio-Roig, Christian Rathgeb, Pawel Drozdowski, Christoph Busch

Abstract—The development of large-scale facial identification systems that provide privacy protection of the enrolled subjects represents an open challenge. In the context of privacy protection, several template protection schemes have been proposed in the past. However, these schemes appear to be unsuitable for indexing (workload reduction) in biometric identification systems. More precisely, they have been utilised in identification systems performing exhaustive searches, thereby leading to degradations of the computational efficiency. In this work, we propose a privacy-preserving face identification system which utilises a Product Quantisation-based hash look-up table for indexing and retrieval of protected face templates. These face templates are protected through fully homomorphic encryption schemes, thereby guaranteeing high privacy protection of the enrolled subjects. For the best configuration, the experimental evaluation carried out over closed-set and open-set settings shows the feasibility of the proposed technique for the use in large-scale facial identification systems: a workload reduction down to 0.1% of a baseline approach performing an exhaustive search is achieved together with a low pre-selection error rate of less than 1%. In terms of biometric performance, a False Negative Identification Rate (FNIR) in range of 0.0% - 0.2% is obtained for practical False Positive Identification Rate (FPIR) values on the FEI and FERET face databases. In addition, our proposal shows competitive performance on unconstrained databases, e.g., the LFW face database. To the best of the authors' knowledge, this is the first work presenting a competitive privacy-preserving workload reduction scheme which performs template comparisons in the encrypted domain.

Index Terms—Biometrics, face recognition, identification, workload-reduction, hashing, secure indexing, privacy protection, fully homomorphic encryption.

1 Introduction

Biometric systems have been successfully deployed in numerous applications such as border control [1]–[3], national identity management systems [4], [5], and forensic investigations [6], [7], among others. According to the International Civil Aviation Organization (ICAO), face recognition has been recognised as one of the biometric technologies most suitable for many practical tasks (*e.g.*, travel documents, national IDs) [8].

Depending on the application context, biometric systems can typically operate in two modes [9], [10]: *verification* and *identification*. Biometric verification is the process of confirming a biometric claim through a one-to-one biometric comparison. In contrast, biometric identification is the process of searching against a biometric enrolment database in order to find and return the biometric reference identifier(s) attributable to a single individual. Generally, a biometric probe is compared against all stored biometric references, thereby leading to a one-to-many biometric comparison (1:N), where N denotes the number of enrolled subjects. In this context, two scenarios can be defined: closed-set identification, which defines that all searched subjects are enrolled in the system, and open-set identification, in which searched subjects are potentially not enrolled in the system. It should be noted, the latter type of scenario (*i.e.*, open-set identification) is more challenging and interesting for real-world applications [11].

Current operational and planned large-scale biometric identification systems around the world (will) host millions or even billions of enrolled subjects [11]. For example, the Indian national ID system, Aadhaar, operates a multi-biometric database of more than 1 billion enrolled subjects [12]. In such large-scale databases, biometric identification based on an exhaustive search is a time-consuming task which is dominated by template comparison costs [4]. In addition, the probability of running into false matches increases with the number of subjects enrolled in the system [13]. It is important to note that an ordering of biometric references is *not* feasible since these do not exhibit an inherent order and are fuzzy [11]. More specifically, different biometric samples, *e.g.*, face images, from the same subject yield templates almost never identical due to biometric variance (*i.e.*, intra-class variations).

Due to the aforementioned issues, numerous researchers proposed different workload reduction (WR) methods for biometric identification systems, which have been surveyed in [11]. The main goal of those approaches is to accelerate the searches in biometric identification systems. In spite of efforts achieved on this topic, most WR approaches still report a degradation of biometric performance while the scalability of some approaches remains questionable [11]. In addition, those schemes usually do not incorporate privacy protection, *i.e.*, biometric references are indexed and retrieved in unprotected form.

Privacy regulations, *e.g.*, the European Union (EU) General Data Protection Regulation 2016/679 (GDPR) [14], usually define biometric information as sensitive data. An unprotected storage of biometric references could lead to different privacy threats such as identity theft, linking across databases, or limited renewability [15]. This has led to the development of Biometric Template Protection (BTP),

• *The authors are with da/sec – Biometrics and Internet Security Research Group at Hochschule Darmstadt, Darmstadt, Germany.*
E-mail: daile.osorio-roig@h-da.de

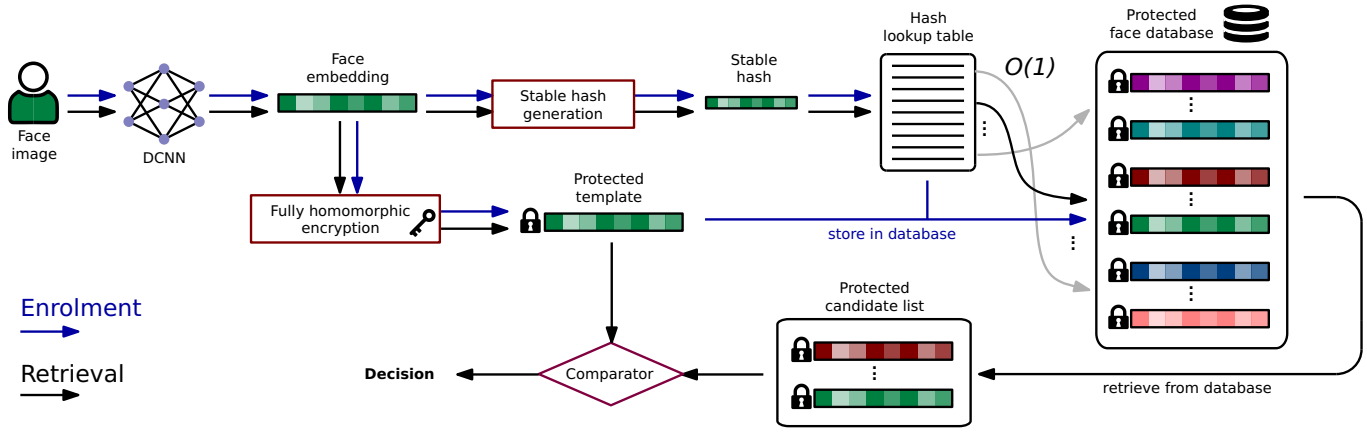


Fig. 1. Overview of a proposed privacy-preserving face identification system: at enrolment a face embedding is extracted from a face image from which a hash and a protected template are generated; subsequently, the hash is used to create an entry in the hash look-up table and the protected template is stored in the database accordingly. At retrieval a hash is generated from a face embedding based on which a subset of protected face templates is retrieved via a hash look-up table; subsequently, the protected probe is compared against the protected candidate list to reach a decision.

i.e., data protection techniques specifically designed for fuzzy biometric data [16]. BTP methods are traditionally categorised as cancelable biometrics, biometric cryptosystems, and homomorphic encryption. BTP schemes must satisfy four main requirements stipulated in the ISO/IEC IS 24745 [17]: *unlinkability*, *irreversibility*, *renewability*, and *performance preservation*. BTP approaches enable a biometric comparison in the protected domain, *i.e.*, biometric templates are permanently protected. However, for the majority of BTP methods, comparisons in the protected domain turn out to be more costly in terms of computational workload compared to the ones carried out by unprotected systems. Consequently, such BTP schemes are less suitable for large-scale identification systems which perform exhaustive searches.

Whereas BTP techniques provide privacy protection, these are hardly employed in biometric identification systems [18]. According to Drozdowski *et al*. [11], only a few works have combined WR strategies with BTP schemes to build large-scale identification systems that ensure privacy protection. In the context of face biometrics, those studies have mainly employed cancelable biometrics [19]–[21]. However, most of those systems still report a degradation w.r.t. biometric performance in benchmark against unprotected systems and may unveil privacy or security issues.

In order to overcome the aforementioned limitations, *i.e.*, provide privacy protection and preserve performance, the use of Fully Homomorphic Encryption (FHE) for face identification was suggested in [22]. This technique, unlike other traditional BTP approaches, preserves biometric performance while the biometric comparison is carried out in the encrypted domain [23]. So far, a few works have employed FHE to achieve privacy-preserving face identification systems, *e.g.*, [22], [24]. In spite of the results obtained by those studies in terms of privacy protection, these systems still perform an exhaustive search to retrieve the protected face references, thereby leading to a degradation of computational efficiency and in an increase of the false match probability depending on the number of subjects enrolled in the system.

Motivated by the aforementioned issues, we propose in this work a face identification system which combines a preselection-based WR strategy with a FHE scheme to fulfil the requirements of ISO/IEC IS 24745 [17] regarding privacy protection. The key contributions of this work are:

- A hash generation scheme based on a Product Quantisation [26] which generates stable hash codes from faces. These hashes are used for indexing a face database, *i.e.*, to construct a hash look-up table. Facial references within the database are protected through FHE. At the time of authentication, face hashes are employed to speed up the retrieval, *i.e.*, to return a candidate short-list. In contrast to existing works in field, the retrieval of the candidate short-list does *not* require a one-to-many search, but can be directly obtained via the hash look-up table, *i.e.*, exact matching with computational complexity of $O(1)$. This is possible since obtained hash codes are highly stable, which further allows for a protection thereof using conventional cryptographic methods. Finally, FHE-based comparisons are carried out in the protected domain for a small fraction of facial references. Thereby, the proposed approach which is depicted in Fig. 1 drastically reduces the overall computational workload of a face-based identification system while the indexing and retrieval is done in a privacy-preserving way.
- A thorough analysis of several clustering techniques to obtain a stable hash generation scheme. The experimental results show the capability of graph- and density-based clustering algorithms to build a stable and compact hash code which can be successfully employed for face identification. In addition, the search of different sub-spaces offered by the PQ- and clustering-based combination allows achieving a good trade-off between efficiency and biometric performance. Moreover, a detailed discussion on the protection of generated hash codes with conventional cryptographic methods is given.
- A literature survey on existing approaches which combine WR strategies with BTP for face identification.

TABLE 1
Overview of most relevant privacy-preserving WR schemes for face-based identification systems (results reported for best configurations and scenarios).

Approach	WR category	BTP category	Dataset	Biometric performance	Exhaustive search
Wang *et al.* [25]	Pre-selection, Feature transformation	Non-traditional BTP	FERET LFW	89% H-R 95% H-R	Yes
Murakami *et al.* [20]	Feature transformation	Cancelable biometrics	NIST BSSR1 SET3	0.1% FRR, 0.022% FAR	Yes
Dong *et al.* [19]	Feature transformation	Cancelable biometrics	LFW (closed-set) LFW (open-set) VGG2 (closed-set) VGG2 (open-set) IJB-C (closed-set) IJB-C (open-set)	99.75% R-1 97.99% DIR, 1% FAR 99.03% R-1 96.03% DIR, 1% FAR 80.57% R-1 56.80% DIR, 1% FAR	Yes
Sardar *et al.* [21]	Feature transformation	Cancelable biometrics	CASIA-V5 IITK CVL FERET	99.85% CRR-1 100% CRR-1 100% CRR-1 100% CRR-1	Yes
Drozdowski *et al.* [22]	Feature transformation	FHE	FERET	$\sim$5% FNIR, 1% FPIR	Yes
Engelsma *et al.* [24]	Feature transformation	FHE	MegaFace	81.4% R-1	Yes
Ours	Pre-selection	FHE	FEI FERET LFW	0.0% FPIR, 0.0% FNIR 0.0% FPIR, 0.2% FNIR 1.0% FPIR, 2.5% FNIR	No

H-R: Hit Rate
FRR: False Rejection Rate
FAR: False Acceptance Rate
R-1: Rank-1 Identification Rate
DIR: Detection and Identification Rate
CRR: Correct Recognition Rate at Rank-1
FPIR:False Positive Identification Rates
FNIR:False Negative Identification Rates

- A comprehensive performance evaluation based on standardised metrics [27] carried out over challenging closed-set and open-set scenarios on three public face databases, *i.e.*, FEI [28], FERET [29], and LFW [30].

The remainder of this work is organised as follows: related works are revisited in Sect. 2. In Sect. 3, the proposed system is described in detail. Sect. 4 presents the experimental evaluations. Finally, in Sect. 5, conclusions are drawn and future works are discussed.

2 Related works

As previously mentioned, numerous efforts have been made in the recent years to avoid an exhaustive search in biometric identification systems. For a summary of state-of-the-art techniques for WR the reader is referred to [11]. According to Drozdowski *et al*. [11], WR methods can be categorised as pre-selection and feature transformation approaches.

Pre-selection algorithms (*e.g.*, data-structures [31], binning [32], and pre-filtering [33]) are focused on the search space reduction, thereby leading to a low number of comparisons (*i.e.*, low penetration rate[1]) per biometric identification transaction. In contrast, feature transformation approaches decrease the computational cost of the individual template comparisons by applying techniques such as reduction of the biometric template dimensionality [34], use of efficient comparators [35], or alignment process [36]. Besides a reduction in workload, feature transformation methods still perform an exhaustive search to determine the identity of a subject. In addition, most WR approaches do not guarantee privacy protection, *i.e.*, only a small amount of works on WR have addressed privacy protection through BTP schemes. Most relevant works are listed in Tab. 1.

Wang *et al*. [25] proposed an obfuscated distance measure which allows concealing the Hamming distance in a dynamic interval. Thereby, privacy protection should be guaranteed while performance is preserved. To that end, the authors proposed a new mechanism to compute earlier distances in the encrypted domain (*i.e.*, Montgomery multiplication domains). This mechanism allows obfuscating the comparison between binary codes-based representations. Then, a collision process of substrings is performed. In this context, hash table-based indexing schemes are built without preserving original distance values. It is important to note that the retrieval of reference templates requires exhaustively comparing pairwise distances in the encrypted domain (*i.e.*, indexing in Montgomery domains).

Murakami *et al*. [20] proposed an indexing scheme to compare so-called *secure indexes (or templates)*. Focusing on privacy, the authors showed that those indexes fulfilled the perfect secrecy property, *i.e.*, transformed indexes leak no information about the original index, *i.e.*, unprotected template. A comprehensive analysis revealed that the proposal complies with the properties of irreversibility, unlinkability, and revocability. In order to build the secure indexes, face templates are transformed by using a permutation process, thereby a cancelable indexing scheme is built. Furthermore, the authors introduced a new method to generate indexes

1. Average proportion of the total number of references that is pre-selected from a database.

from biometric features through Generative Adversarial Network (GAN) discriminators which they refer to as *pivot*. The pivot-based strategy basically produces a distortion over the features of subjects depending on how their pivots differentiate from other subjects. That is, this strategy allows generating a different distribution of the unprotected biometric features.

Recently, Dong *et al*. [19] proposed an identification system where privacy was ensured by transforming facial features to more compact non-invertible features. As a consequence, faster comparison (1:1) in the Hamming space could be achieved. To this end, an Index-of-Maximum (IoM) [37] locality sensitive hashing-based technique [38] was used. In general, BTP schemes based on cancelable biometrics were reported to negatively affect biometric performance [16].

In contrast, the approach in [19] achieved competitive results with respect to its baseline (*i.e.*, unprotected system) by employing different fusion strategies. Inspired by this idea, Sardar *et al*. [21] introduced a novel hashing technique, namely *FaceHashing*. Within this method, privacy protection is achieved by modifying the BioHashing technique [39], [40]. In particular, this approach computed a cancelable feature vector by applying several feature transformations (*i.e.*, sparse representation, coordinate descent, and block coordinate descent techniques).

It is worth noting that the aforementioned research efforts usually yield a trade-off between privacy protection, computational efficiency, and biometric performance within face identification systems. Some of these works still suffer from significant performance degradations (*e.g.*, [19]). Other works do not provide sufficient privacy protection, *i.e.*, they leak biometric feature information (*e.g.*, [41], [42]). In addition, not all methods have properly shown, that they can reach unlinkability for cross-comparison attacks [15]. Further, WR strategies employed by these systems (*e.g.*, [19]–[21]) are limited to feature transformation or exhaustive pre-selection approaches (*e.g.*, [25]).

In order to overcome the aforementioned issues, some face identification systems have recently introduced FHE-based BTP schemes. Engelsma *et al*. [24] proposed a compact feature representation through intrinsic dimensionality reduction based on Deep Neural Networks [43]. In addition, the authors introduced a Brakerski/Fan-Vercauteren scheme-based [44] encoding strategy. Keeping in mind that operations in the encrypted domain are expensive, the authors proposed to encode each intrinsic dimension of the compact feature as a matrix in order to decrease the computational complexity. Technically, a set of plaintexts (*i.e.*, set of dimensionality-reduced feature vectors) are encoded in a single ciphertext. Finally, an inner product between two ciphertexts in the encrypted domain could be carried out efficiently in the comparison stage. In spite of the remarkable results, the used encoding scheme is restricted by the number of compressed biometric templates and encryption parameters, thereby resulting in a performance degradation.

Drozdowski *et al*. [22] proposed a simpler biometric face identification system in which privacy protection is achieved by the application of FHE. In their work, the authors evaluated two encoding schemes (*i.e.*, Cheon-Kim-Kim-Song [45] (henceforth referred to as "CKKS") and Brakerski/Fan-Vercauteren [44] (henceforth referred to as "BFV")) in a trusted third party-based architecture. In this case, the encoding process is carried out for each feature vector, in contrast to the method proposed in [24]. It is concluded that there exist several challenges and issues which must be dealt with for such schemes, *e.g.*, FHE-based BTP schemes, to be viable in a biometric identification scenario, especially if WR is to be employed [22]. Some of these challenges can be summarised here:

- The speed up of FHE-based scheme implementations may be prohibitive for larger deployments. In this context, concepts of WR could be introduced in order to narrow down the search space for each biometric identification transaction.
- One drawback of using HE is that it limits the flexibility in the implementation. For instance, feature vector elements may not be accessible individually. This limitation makes incremental recognition schemes (which facilitate early acceptance/rejection of likely/unlikely candidates) infeasible.
- The incorporation of search strategies, *e.g.*, binning, indexing, or 1-to-first-based search, over these BTP schemes may lead to a trade-off between biometric performance, computational workload, as well as data security and privacy in a biometric identification system.

3 Proposed System

The proposed scheme consists of four main steps: at the time of enrolment, $i)$ a reference face image (*i.e.*, input) is captured, a face is detected, pre-processed, and its feature representation, denoted as S, is extracted (Sect. 3.1); $ii)$ for each S, the hash generation scheme extracts a hash code, $H(S)$, which is stored as an index in a hash look-up table; $iii)$ additionally, S is encrypted (*i.e.*, $Enc(S)$) through the BFV encoding scheme [44] which is used as base in the FHE scheme [46]. In the enrolment, $iv)$ $Enc(S)$ is stored as protected reference template in its corresponding index (*i.e.*, generated hash code) in the hash look-up table (Sect. 3.3). At the time of authentication, a probe face image is captured, processed in the same way (following the steps $i)$, $ii)$, and $iii)$). Subsequently, $iv)$ a hash code is retrieved from the hash look-up table. Finally, a candidate score list of protected references stored at this hash code and the protected probe template is compared against them.

In our work, inspired by the security protocol in [24], we adopt a semi-honest model where each party (*i.e.*, entity) is constrained by the following protocol [47]. In particular, our system is built upon three entities: $i)$ a client, which provides the biometric face features for enrolment or authentication: at the time of enrolment, the client supplies the public key to encrypt the reference face features, while at the time of authentication, the client provides the private key to encrypt the probe face features and decrypt the scores computed; $ii)$ a database, which holds a hash look-up table under the responsibility of storing the encrypted reference templates; and $iii)$ a server being the channel of communication between the client and the database; both the enrolment and authentication are performed in the encrypted domain. It is worth noting, that in the context of authentication, the encrypted candidate list containing the

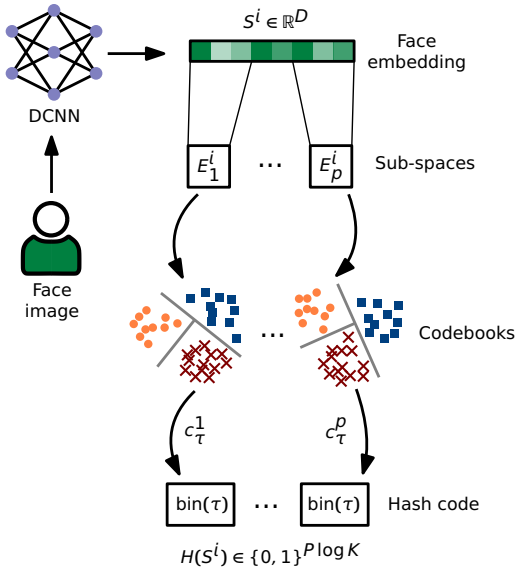


Fig. 2. Overview of the proposed hash generation scheme.

scores of the reference templates most similar to the subject at hand is directly transmitted from the server to the client party. By using its private key, the client finally decrypts the scores of the candidate list and the best score (*i.e.*, after sorting the list) is selected as final biometric comparison score.

3.1 Feature representation

The client party supplies the biometric features of an input face image. To that end, embedding representations extracted from several Deep Convolutional Neural Networks (DCNN)-based face recognition systems are used. Therefore, in the first step, faces are detected and aligned by using the Multi-task CNN [48] framework. Then, the face embedding is extracted. In our experiments, three well-known and open-source face recognition systems are evaluated: FaceNet [49], ArcFace [50], and VGGFace2 [51], which have shown a remarkable performance for face recognition tasks.

3.2 Hash generation

An overview of the proposed hash generation scheme is depicted in Fig. 2. In order to extract stable hash codes from faces, we adopt the Product Quantisation (PQ) approach [26]. This technique yields a compact discrete representation from data which can be employed for either exhaustive or inverted indexing searches [52]. In particular, in our work, we explore PQ in combination with clustering techniques to generate a stable binary hash code which, in turn, allows a fast retrieval from the hash look-up table since it enables an exact (non-fuzzy) comparison like it is the case for passwords or PINs. This means, for a probe face, a candidate short-list is retrieved with a computational complexity $O(1)$. It is worth noting that such clustering techniques have been successfully applied to face images [53].

Face embeddings of size D are extracted from N subjects. Note that more than one face image per subject could be used in the enrolment stage. Let $\mathbf{S} = \{S^1, \ldots, S^N\}$ be a set of face embeddings of all subjects to be enrolled. Each face embedding can be represented as a concatenation of P sub-spaces, *i.e.*, a set of equal-size sub-vectors that constitute the face embedding, each of dimension $\frac{D}{P}$, denoted as $S^i = \{E_1^i, \ldots, E_P^i\}$.

For each sub-space $1 \leq j \leq P$, the PQ generates a codebook $C^j = \{c_1^j, \ldots, c_K^j\}$ of size K, where K represents the number of codewords (or number of clusters) in C^j. Subsequently, a sub-vector E_j^i is mapped to a codeword c_τ^j in its corresponding codebook C^j. The value τ indicates the index of the nearest codeword c_τ^j to the sub-vector E_j^i and can be represented as binary hash code (bin) with $\log_2(K)$ bits. In summary, a binary hash code $Q(E_j^i)$ is obtained for a sub-vector E_j^i of the face embedding S^i:

$$Q(E_j^i) = \underset{\text{dist}(E_j^i, c_\tau^j)_{\tau=1,\ldots,K}}{\text{argmin}} \text{bin}(\tau), \tag{1}$$

where $\text{dist}(E_j^i, C_\tau^j)$ is the similarity between E_j^i and the nearest codeword $c_\tau^j \in C^j$. Finally, the hash code of size $P \log_2(K)$ bits representing a face embedding S^i is estimated by concatenating the binary hash code $Q(E_j^i)$ for each E_j^i:

$$H(S^i) = \sqcup Q(E_j^i) : j = 1, \ldots, P \tag{2}$$

As aforementioned, PQ builds a set of codebooks from data. In our work, we evaluate four different clustering algorithms (*i.e.*, K-means [54], K-medoids [55], Gaussian mixture models (*i.e.*, GMM) [56], and Affinity propagation (*i.e.*, AP) [57]) to generate those codebooks.

3.2.1 K-means

K-means is a well-known centroid-based clustering technique which yields a partition of N observations into k clusters defined a priori [54]. In particular, data points are assigned to k groups by minimizing the squared error distance between them. Each cluster computed by K-means is represented by a fictitious node, the so-called centroid (average of all the points in its cluster). Therefore, to assign a new point to a cluster, K-means calculates its distance with the closest centroids. The K-means computational complexity is $O(kN)$ and due to its rapid convergence, this clustering algorithm has been widely used in numerous computer vision and pattern recognition tasks [58]–[60]. In our work, the centroids represent a codebook for a particular sub-space.

3.2.2 K-medoids

K-medoids [55] is a centroid-based clustering technique similar to K-means, which assumes k clusters a priori. In contrast to K-means, this technique computes its centroid in a different way, *i.e.*, through its medoid. Medoids are the most centrally located data points in the clusters, with the minimum sum of distances to other points. It is important to note that medoids are always restricted to be members of the data set. In our work, these data points represent a codebook which is generated for each sub-space.

3.2.3 Gaussian mixture models

GMM [56] is a density-based clustering algorithm which assumes each point stems from a mixture of Gaussians. Like earlier centroid-based approaches, GMM needs to define the number of clusters K beforehand. In particular, GMM introduces a degree of dependence or uncertainty between the partitions by assigning probabilities (*i.e.*, soft assignment). In this context, GMM can be understood as a probabilistic visual vocabulary whose clusters are described by their mixture weights, means, and covariance matrices. In our work, for Product Quantisation, we compute the loglikelihood between a sub-vector and the corresponding probabilistic GMM codeword to look for the most appropriate cluster.

3.2.4 Affinity propagation

In contrast to the aforementioned clustering algorithms, Affinity propagation (AP) [57] does not require a number of clusters K a priori. The AP recursively transmits real-valued messages along the edges of a graph until a good set of *exemplars* and corresponding clusters emerge. To that end, AP takes as input a set of real-valued similarities between data points, where the similarity $s(i, j)$ indicates how well a data point with index j is suited to be a possible exemplar for the data point i. In order to find the best exemplar, AP minimises squared errors, where each similarity is set to the Euclidean distance. In our pipeline, a sub-vector is assigned to one of M exemplars representing the clusters in the codebook.

3.3 Template encryption

FHE schemes provide privacy protection of the stored template by computing comparisons between them in the encrypted domain. In addition, it provides a high biometric performance since biometric comparators (*e.g.*, Euclidean distance) can be computed without a negative impact in the homomorphic domain. In particular, we adopt BFV scheme [44] as the base for the FHE computation. The BFV scheme has shown a significant speed-up in the encrypted domain computation, in contrast to CKKS scheme [22]. Also, the BFV scheme allows the use of the packing or batching technique [61] which encrypts multiple values into a single ciphertext (*i.e.*, facilitates operations on vectors component-wise) and hence performs computations by using the SIMD (Single Instruction Multiple Data) primitives [62], [63]. Keeping in mind that the BFV schemes operate on integer values [44], face representations are firstly quantisised following the equal-width quantile strategy [64] which assigns for each float-value component in the vector an integer value.

A single one-to-one comparison based on BFV between an encrypted probe and each reference template stored in the candidate list is based on the aforementioned batching technique. Specifically, we compute the squared Euclidean distance between two vectors in the encrypted domain. In particular, addition operations are applied by cyclically rotating the vectors without the need of decrypting them [63].

3.4 Hash look-up table

In order to speed up the subject retrieval, we make use of a hash look-up table. This hash look-up table stores the hash code $H(S^i)$ generated by PQ and as entries, their corresponding encrypted templates (*i.e.*, $Enc(S^i)$), hence a candidate list can be returned with a computational complexity of $O(1)$. In this context, it is expected that the same hash code $H(S^i)$ may point at various encrypted templates $Enc(S^i)$. In a nutshell, the same hash code $H(S^i)$ could point at several encrypted templates of different subjects. Given a hash code from a probe, our pipeline allows the retrieval of a candidate short-list with size $t \ll N$ which contains the encrypted reference templates of the subjects most similar to the probe. All encrypted templates in the candidate list returned are compared with the encrypted probe. Finally, a list of encrypted scores is obtained. For a non-existent hash code value, an empty candidate list is returned.

To prevent from cross-matching or reconstruction attacks, it is recommended to encrypt the obtained hash code with conventional cryptographic methods. More specifically, it is suggested to employ an application-specific key. For instance, this could be realised with Message Authentication Codes (MACs) which may involve the use of cryptographic hash functions (HMACs), *e.g.*, SHA-256. However, MACs usually produce bitstrings which are expected to be significantly longer than the hash code extracted in the proposed method. Therefore, it is suggested to only use the first $P \log K$ bits (length of hash code) returned by the MAC. Thereby, the size of the hash look-up is maintained.

3.5 Workload reduction

In this section, we define the cost of a biometric transaction in the proposed privacy-preserving identification system. To that end, we analyse the workload (W) of our system in a single lookup. We compute W as follows:

$$W = N \times p \times \theta + \beta \tag{3}$$

with:

- N: number of enrolled references
- p: penetration rate
- θ: computational cost of a single one-to-one comparison in the encrypted domain by using the BFV-based encoding scheme. It should be noted that our system takes 750ms over the BFV-based encoding scheme.
- β: computational cost when encrypted templates are retrieved through the pre-selection-based WR strategy, *i.e.*, the hash look-up table. Note that our system enables an exact match which has a computational complexity of $O(1)$ and hence a retrieval low computational cost, in contrast to other pre-selection methods, *e.g.*, Wang *et al*. [25]. In this context, our system takes 0.003ms to pick up a small fraction of protected templates.

In our system, p is the average proportion of retrieved or pre-selected candidates for each hash code in the hash look-up table. It is determined as follows:

$$p = \frac{\gamma}{N} \tag{4}$$

where γ represents the average number of comparisons per hash code when a lookup is carried out, and N is the number of subjects, *i.e.*, reference templates stored in the system database. It is worth noting that p depends

TABLE 2
Summary of the face embeddings extracted from different face recognition systems.

Face recognition system	Pre-trained model	Feature embedding size	Loss function
FaceNet	Inception-ResNet-v1[1]	512	Triplet
ArcFace1	ResNet-100[2]	512	Additive Angular Margin
ArcFace2	MobileFaceNet[3]	128	Additive Angular Margin
VGG-Face2	Senet-50[4]	2048	Soft-max

[1] https://github.com/davidsandberg/facenet
[2] https://github.com/deepinsight/insightface/wiki/Model-Zoo
[3] https://github.com/deepinsight/insightface/wiki/Model-Zoo
[4] https://github.com/ox-vgg/vgg_face2

on the number of used sub-spaces P and codewords K. The maximum number of hash codes computed by $H(\cdot)$ (see equation 2) can be represented as K^P. Therefore, the probability of collision (*i.e.*, f) can be determined as follows:

$$f = \begin{cases} 1 & \text{if } N > K^P \\ \frac{N}{K^P} & \text{otherwise.} \end{cases}$$

As it should be noted, if N is a constant value and P increases, then f decreases and hence a low p and consequently a low W is achieved.

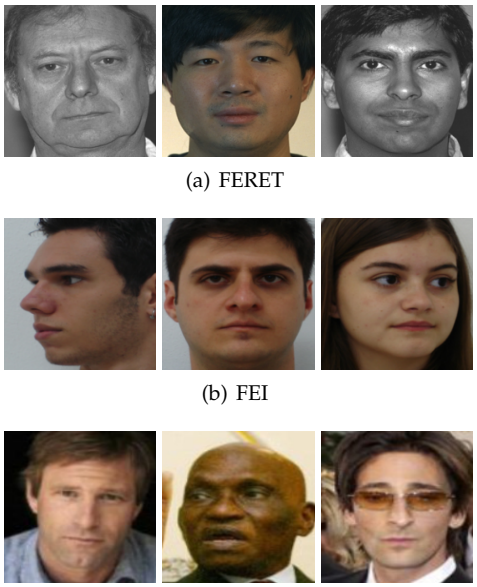
4 Experimental Evaluation

In this section, we evaluate the performance of the proposed system using a hash look-up table constructed with PQ in combination with different clustering methods. To that end, three goals are defined: i) analyse the biometric performance for several parameter configurations over closed-set scenarios ii) evaluate the trade-off between biometric performance, privacy protection, and workload reduction for open-set scenarios, and iii) a benchmark with other state-of-the-art systems.

4.1 Experimental Protocol

Three different face recognition systems (*i.e.*, FaceNet, ArcFace, and VGG-Face2) are selected to extract face embeddings. In our evaluation, pre-trained models provided for each face recognition system are used. In particular, two pre-trained models of ArcFace (hereafter referred to ArcFace1 and ArcFace2 over each pre-trained model) are employed. Tab. 2 shows a summary about the systems utilised in our work. Our baseline is an exhaustive search, *i.e.*, a biometric probe is compared against all references enrolled in the database exhaustively. The proposed identification system is fully implemented in Python. The Scikit-Learn library[2] is used for the computation of different methods employed in our investigation (*i.e.*, K-means, K-medoids, GMM, and AP). In addition, a PySeal wrapper[3] on Python 3.7, which uses the C++ SEAL open-source library [65] is utilised for FHE. Focusing on the encryption parameters, we select those parameters that corresponded to a security level of 128 bits[4]. Higher levels of security enabled by the library, *e.g.*, 192 and 256 bits, lead to higher execution times [66]. Although it is true that most homomorphic encryption schemes provide weaker security guarantees than traditional encryption schemes, it is not the actual limiting factor in the case of facial biometrics, as discussed in section 4.2.2.3. In our system, we guarantee a trade-off between efficiency and security for a security level of 128 bits. Our evaluations were conducted on an Intel Core i7-8750H@2.2GHz, 16GB RAM hardware, Linux environment.

2. https://scikit-learn.org/stable/
3. https://github.com/Lab41/PySEAL
4. According to the https://homomorphicencryption.org/ standard.



(a) FERET

(b) FEI

(c) LFW

Fig. 3. Example images from the selected datasets (a) FERET (b) FEI (c) LFW.

4.1.1 Datasets

Three publicly available datasets, as shown in Fig. 3, are used in our investigation:

- **FERET** [29] includes 14,126 facial images from 1,199 subjects. In our experiments, we select individuals with frontal images without intentional occlusions (*e.g.*, scarves or sunglasses). Some facial images with exceedingly poor quality are also removed, as done in [67]. Finally, the selected subset comprises 2,697 face images from 987 subjects where the first subset of them comprises subjects who contain at least three samples, while a second subset contains subjects having two

samples. Given that these subsets are unbalanced, we defined protocols for the investigated scenarios (*i.e.*, closed-set and open-set scenarios):

- For the close-set scenario, we select the subset having more than three samples per subject. For training and enrolment, three samples are randomly picked up while the remaining images are included in the search.
- For the open-set scenario, we select the same set of images in the closed-set for training, enrolment, and search. In addition, the subset having two samples per subject is included in the search set.

- **FEI** [28] consists of 2,800 color images from 200 subjects. Most subjects contain 14 images with pose rotation up to 180 degrees. Some images, whose faces (*e.g.*, p13 and p14) are not visible, were removed from the dataset. For the closed-set evaluation, ten samples per subject are randomly selected for training and enrolment while the remaining samples are used as probe samples. Note that a 5-fold cross-validation is considered since FEI does not present a defined protocol for open-set scenario, in contrast to LFW. Each of the search sets contains approximately 136 samples from different subjects. In addition, it should noted that we found an labelling error (*i.e.*, samples with labels 72 and 2 stem from the same subject) on the FEI database. This error was corrected on open-set scenario, thereby leading to 199 subjects for FEI.
- **LFW** [30] is the first dataset focused on the large-scale unconstrained face recognition problem. This comprises 13,233 face images from 5,749 subjects collected from the web where 1,680 subjects are represented with two or more images. While the remaining subjects have a single image. We divide LFW into three different sets as done in [19]:
  - **Known** subject set contains 6,733 face images from 610 subjects having more than three samples. These images are only included in the enrolment and training set.
  - **Known unknown** subject set consists of 2,431 face images from 1,070 subjects who are used for training our hash generation scheme, yet not for enrolment. Those subjects have two or three samples.
  - **Unknown unknown** subject set comprises 4,069 face images from 4,069 subjects which are not used for training either in the enrolment set. They only have a single sample.

  In the closed-set evaluation, we followed the protocol in [19], where three samples per subjects in **Known** set and one image per subject in **Know unknown** are randomly selected for training our hash generation scheme. In addition, the three same samples selected per subject from the **Known** set are employed for enrolment while the remaining images are included in the identification as probe faces.
  On the other hand, for open-set evaluation, we followed the two protocols in [19]:
  - **Open-set O1 scenario** contains the images from the closed-set evaluation for training, enrolment, and search. Samples in the **Know unknown** set which are not included in the training are also used for search.
  - **Open-set O2 scenario** contains the same images from the closed-set evaluation for training, enrolment, and search. In addition, the search set includes images of the **Unknown unknown** set.

TABLE 3
Summary of databases, scenarios, and partition into training, enrolment, and search sets.

Dataset	Scenario	Training	Enrolment	Search
LFW	Closed-set	2,898	1,830	4,902 genuines
	Open-set O1	2,898	1,830	4,902 genuines 1,359 impostors
	Open-set O2	2,898	1,830	4,902 genuines 4,069 impostors
FERET	Closed-set	747	747	474 genuines
	Open-set	747	747	474 genuines 1,476 impostors
FEI	Closed-set	2,000	2,000	776 genuines
	Open-set	1,890	1,890	732 genuines 136 impostors

A detailed description of the number of samples for training, enrolment, and search for both scenarios is listed in the Tab. 3. It is worth noting that the images used for the hash generation scheme training as well as the ones in the enrolment are randomly selected. In addition, the 5-fold cross-validation is carried out on both evaluations (*i.e.*, open-set and closed-set).

4.1.2 Metrics

The experimental evaluation is conducted according to ISO/IEC 19795-1 [27] standard methods and metrics:

- Pre-selection error rates are the proportion of subjects for which the corresponding subject identifier is not in the pre- selected subset of candidates.
- Hit-rate, which computes the complement of the pre-selection error rates: 1 - pre-selection error rates.
- Identification rate which is reported as a cumulative match characteristic (CMC) plot. The CMC plots the rank-R identification rate. Rank-1 is considered in our evaluations.
- False Negative Identification Rates (FNIR), which is defined as the proportion of a specified set of identification transactions by subjects enrolled in the system for which the subject's correct reference identifier is not among those returned.
- False Positive Identification Rates (FPIR), which is defined as the proportion of identification transactions by subjects not enrolled in the system for which a reference identifier is returned.
- Workload (W) which is defined as the overall computational workload of a biometric identification transaction as a percentage of the baseline (exhaustive search) workload in terms of the number of template comparisons. It should noted that in this work W is computed as defined in Sect. 3.5.

TABLE 4
Rank-1 identification rates of the baseline systems.

System	Dataset		
	FEI	FERET	LFW
ArcFace1	100.00 %	100.00 %	99.84 %
ArcFace2	98.97 %	100.00 %	99.67 %
FaceNet	99.74 %	100.00 %	85.08 %
VGGFace	99.82 %	99.79 %	99.84 %

Based on those metrics, we also report: i) the Detection Error Trade-off (DET) curves between FPIR and FNIR and ii) the FNIR observed at different FPIR values such as 1% (FNIR100) and 0.1% (FNIR1000).

4.2 Experimental Results

4.2.1 Closed-set scenario

4.2.1.1 Baseline feature extractor: In the first experiment, we select the most appropriate face embeddings together with their corresponding pre-trained models (see Tab. 2) and carry out identification over closed-set scenarios per database without applying workload reduction efforts (*i.e.*, exhaustive search). Tab. 4 lists the Rank-1 identification rates obtained by the different pre-trained models on all considered databases. All used pre-trained models achieve high identification rates. In particular, the ArcFace recognition system (*i.e.*, ArcFace1) shows the best biometric performance across the used datasets. Therefore, ArcFace1 is considered for the following experiments.

4.2.1.2 Impact of parameters: In the second experiment, we determine the optimal configuration of our hash generation scheme in terms of the main parameters: the number of centers or codewords K and the number of sub-spaces P. To that end, we select the value range $\mathbf{K} = \{64, 128, 256, 512, 1024\}$ and $\mathbf{P} = \{1, 2, 4\}$. Values for K and P which are greater than 1024 and 4, respectively, would result in large binary hash codes which are not suitable for a robust hash generation. The pre-selection error rates per parameter configuration, clustering algorithm, and dataset are summarised in Tab. 5. As it can be observed, the best pre-selection error rates are obtained for different K per database. Among centroid-based clustering methods (*i.e.*, K-means and K-medoids), the former obtains its best average pre-selection error rates at $K = 256$, thereby resulting in an average pre-selection error of 4.8%. In addition, the GMM achieves its best average pre-selection error rate at $K = 256$, which is up to eight times lower than the ones achieved by centroid-based approaches. This indicates that $K = 256$ mixture of Gaussians are enough to successfully decrease the intra-class variability between samples from the same subject and increase the inter-class variability between different subjects. It is important to highlight that, no result is reported for at $K = 1024$ for the FERET database since the number of samples used for hash generation is lower than 1024 (*i.e.*, 747).

Keeping in mind that the AP does not require the number of codewords K a priori, we also report in Tab. 5 its biometric performance for the number of exemplars (*i.e.*, K) detected by this approach. In particular, AP yields an average pre-selection error rate which is approximately three times lower than the one achieved by GMM. It should be noted, the AP obtains its best biometric performance at the different number of centers depending on the dataset at hand (*i.e.*, K = 256 for FEI and FERET, and K = 1024 for LFW). This is due to the intra-class variation and face image quality of samples in the LFW, which are more challenging than the those in FEI and FERET. In contrast to K-means, K-medoids, and GMM, AP depends on the inter-and intra-class variabilities on the dataset to detect reliable exemplars. Given that the AP finds the optimum number of clusters representing the data, the number of hash codes generated by PQ depends on the dataset as well. Challenging databases such as LFW could benefit from this type of clustering algorithm. In order to validate earlier affirmations, a thorough analysis of the effect of image quality on the hash generation is provided. To that end, AP and GMM, which reports the best average pre-selection error rates, are considered. Further, we observe a biometric performance degradation as the number of $\mathbf{P}$ subspaces increases. Specifically for those clustering methods whose number of centers needs to be defined a priori (*i.e.*, K-means, K-medoids, and GMM).

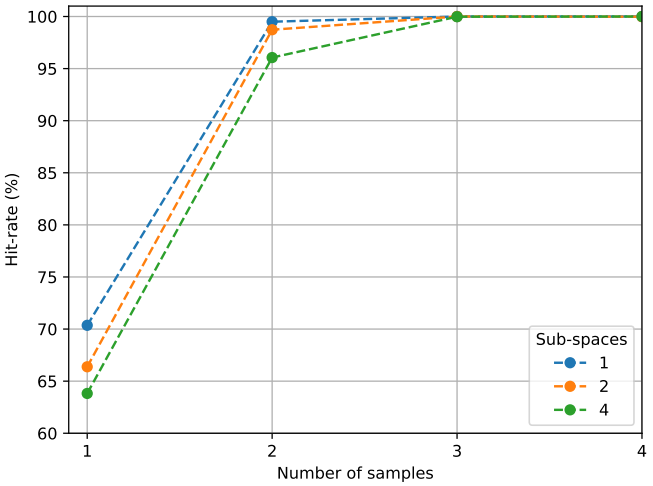


Fig. 4. Analysis of the number of samples per subject used for training the AP-based hash generation.

4.2.1.3 Effect of the number of samples for the hash generation training: In order to analyse the robustness of the proposed hash generation scheme, we explore the effect of the number of samples used for its training. To this end, we use the proposed AP-based hash generation scheme. In addition, the FEI database, which contains the highest number of samples per subject, is chosen. Fig. 4 reports the hit-rates for several numbers of samples used for training the hash generation over different P sub-spaces. In all experiments, we randomly select the number of images used for training. The remaining samples are included in the probe set (*i.e.*, search set). It can be observed that AP achieves a reliable biometric performance by using only three samples for training exhibiting high robustness over a datasets like FEI containing faces with different poses variations.

4.2.1.4 Effect of the face image quality: As could be seen in Tab. 5, the biometric performance on LFW improves

TABLE 5
Biometric performance, in terms of pre-selection error rates (%), of the proposed hash generation scheme for different values of K (centers) and P (sub-spaces). The best results per clustering method are highlighted in bold.

Database	Centers	K-means			K-medoids			GMM			Affinity Propagation		
		$P=1$	$P=2$	$P=4$	$P=1$	$P=2$	$P=4$	$P=1$	$P=2$	$P=4$	$P=1$	$P=2$	$P=4$
FEI	64	**0**	0.05	1.31	6.75	18.74	46.65	**0**	0.41	15.70			
	128	**0**	**0**	0.13	5.00	11.21	30.93	**0**	**0**	2.37			
	256	0.05	0.15	0.98	1.98	5.28	12.73	0.05	0.1	0.54	**0**	0.03	0.05
	512	0.03	2.55	10.52	0.54	3.76	14.59	0.1	1.57	6.24			
	1024	**0**	6.16	17.5	**0.21**	9.36	13.46	0.08	5.21	13.76			
FERET	64	**0**	0.46	2.95	18.48	39.49	74.25	**0**	0.21	2.57			
	128	**0**	0.04	1.77	14.3	68.78	60.13	**0**	0.25	0.72			
	256	**0**	**0**	0	11.90	20.76	39.28	**0**	**0**	0.08	**0**	**0**	**0**
	512	**0**	5.49	13.38	**5.86**	12.28	32.15	**0**	4.09	14.3			
LFW	64	7.01	18.43	44.19	32.82	57.46	77.23	3.88	8.08	36.32			
	128	6.43	15.48	33.73	33.08	61.03	77.19	4.19	7.76	21.79			
	256	5.01	10.53	26.69	27.51	58.02	76.89	3.59	6.91	16.34	2.28	**2.23**	3.14
	512	3.24	5.31	10.36	21.11	48.51	73.39	3.37	8.59	5.09			
	1024	**2.19**	**2.19**	2.45	**18.51**	33.5	61.9	**2.14**	2.26	3.09			

with K for each clustering-based hash generation, in contrast to FEI and FERET. Even, for the AP method which does not require a number of clusters a priori, it can be observed that for LFW lowest error rates are achieved for $K = 1024$. This is because the samples per subjects in LFW tend to exhibit higher variations in contrast to images in FEI and FERET. This is also reflected in the baseline performance rates of Tab. 4.

Based on that fact, we explore the impact of image quality on the hash generation used by our identification system. To that end, we first compute on the LFW, FEI, and FERET databases, in Fig. 5, the scores of image quality for all their samples through FaceQNet [68]. As it may be observed in Fig. 5, LFW exhibits the highest variation of image-quality scores ranging from 0.0 to 0.7. More precisely, the amount of samples with lower scores, *e.g.*, < 0.3, is greater than the ones reported by FEI and FERET. Note that the performance of the hash generation scheme could be sensitive to the image-quality variation, thereby confirming the results on Tab. 5 for small K values. Also, it should be noted that low quality scores in the FEI database mainly result from pose variations while LFW contains many low resolution face images. Secondly, we calculate, in Fig. 6, the biometric performance for the GMM-based hash generation scheme over different K values and five image quality thresholds. It can be observed that the hit-rates improve with K for those ranges including poor quality images. As expected, the robustness of the proposed hash generation scheme improves with image quality.

Fig. 7 depicts a probe and references from LFW which resulted in false matches in the proposed identification system. Both, the probe and the references exhibit a rather low face image quality. According to Shi and Jain [69], distances of face embeddings in the latent space can be distorted for low quality sample pairs, thereby leading to false matches.

4.2.1.5 Benchmark with the baseline: In this experiment, the best configurations of the GMM- and AP-based system are compared with the baseline (*i.e.*, ArcFace1). From the results in Tab. 6, only slight deterioration in biometric performances can be observed in comparison to the baseline. However, it can also be observed that some configurations maintain the biometric performance of the baseline system on specific datasets. For instance, for the GMM-based system using $P = 4$ sub-spaces an identification rate of 98.87% is achieved on the LFW database while this configuration yields a slight decrease of biometric performance on the FEI dataset.

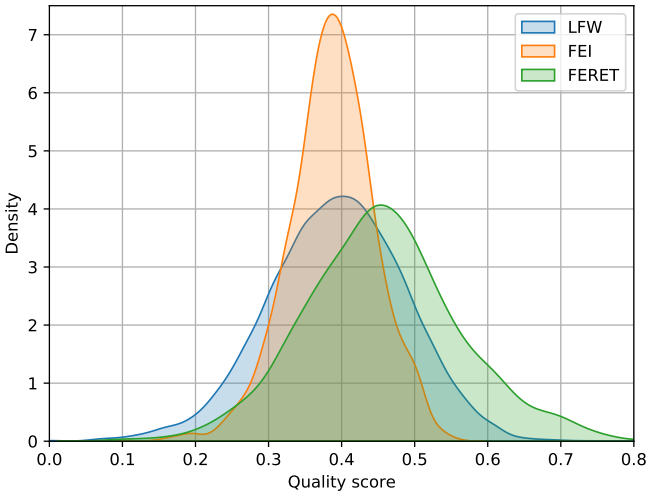


Fig. 5. Probability density function of quality scores computed by FaceQNet for the LFW, FEI, and FERET databases.

TABLE 6
Benchmark of our hash generation scheme for different P values with the ArcFace1 baseline in terms of identification rate (%) at rank-1.

Dataset	GMM			AP			Baseline
	$P=1$	$P=2$	$P=4$	$P=1$	$P=2$	$P=4$	
FEI	100.00	99.98	99.38	100.00	100.00	100.00	100.00
FERET	100.00	100.00	100.00	100.00	100.00	100.00	100.00
LFW	99.67	99.81	98.87	99.84	99.68	99.68	99.84

4.2.2 *Open-set scenario*

4.2.2.1 Biometric performance: As it was mentioned, one of the goals of the proposed system is to correctly handle identification transactions with data subjects whose references are not present in the enrolment database (*i.e.*, open-set scenario). To that end, we consider the eval-

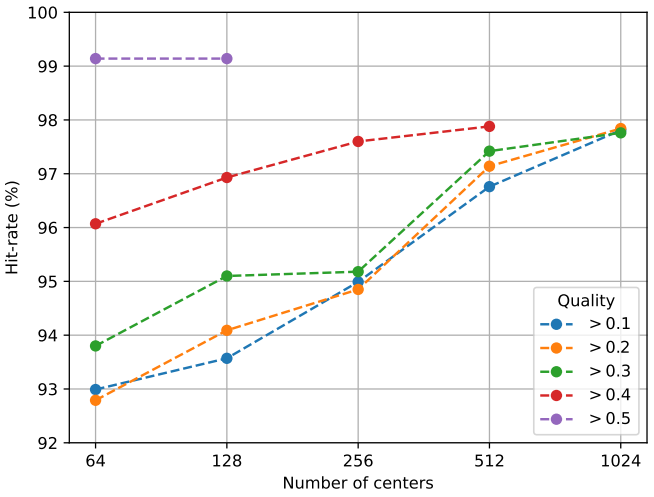


Fig. 6. Impact of the quality over the LFW database in terms of hit-rate for the GMM-based scheme using K centers and $P = 1$ (for higher quality thresholds the number of samples becomes too low to obtain meaningful results).

(a) probe: quality = 0.38

(b) references: quality < 0.44

Fig. 7. Example of false match between a probe and references.

uation of our best performing hash generation scheme (*i.e.*, AP) to retrieve non-encrypted and encrypted face templates from the selected databases.

For the unprotected as well as the protected system, on the FERET dataset a low FNIR of approximately 0.2% is achieved at a FPIR = 0.0%. Similarly, on the FEI database a FNIR = 0.0% is obtained for a FPIR = 0.0% (*i.e.*, perfect separation). In general, the performance of the unprotected system is maintained. Slight performance variations are caused by the feature quantisation step which is necessary for the applied encoding scheme. Future work could be focused on a float-based encoding scheme, *e.g.*, CKKS [45], in combination with feature transformation-based WR strategies to improve its efficiency. On the other hand, on the LFW database, at a FPIR = 1.0% lowest FNIR of approximately 2.5% are achieved for using $P = 1$ and $P = 2$ subspaces. Corresponding DET curves for the different scenarios of the LFW dataset are plotted in Fig. 8 (DETs for FEI and FERET are omitted due to their low rates and hence these do not contain data points obtained from a significant amount of identification attempts).

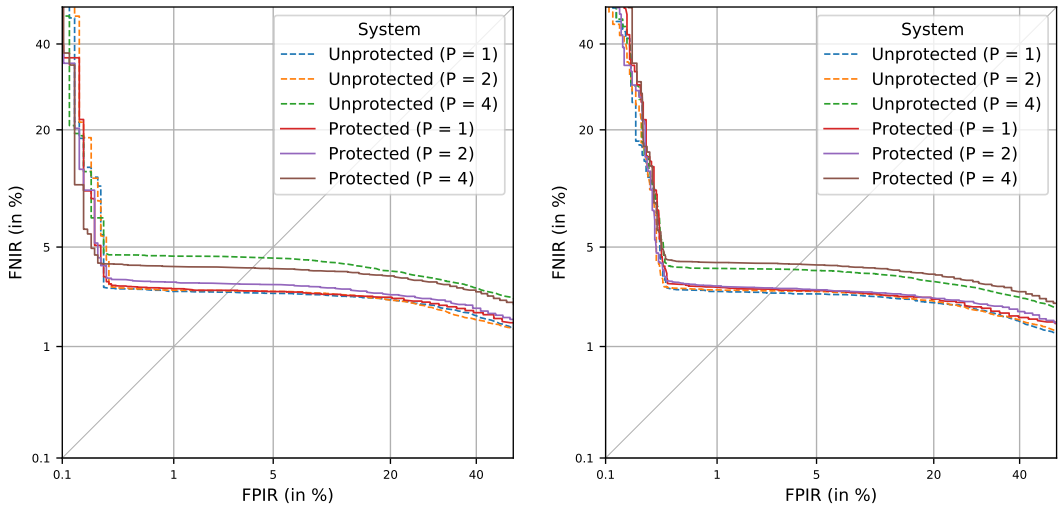
Based on the obtained results, it should be noted that biometric performance may decrease in open-set scenarios (compared to closed-set), mainly depending on the database characteristics. Note that the best biometric performance is obtained on less challenging databases, *e.g.*, FEI and FERET, in contrast to LFW. On LFW, a significant drop in biometric performance is observed, compared to the closed-set scenario. This can be observed by comparing the corresponding results in Tab. 4 to those in Fig. 8. Even for identification over closed-set scenarios, variations in identification rates for the different databases can be seen. Additionally, Sect. 4.2.1.4 confirms the differences over those databases in terms of image quality. In case face images exhibit generally lower quality, false positives can occur with higher probability since extracted face embeddings tend to be less discriminative. Consequently, identification transactions of impostors which are not enrolled in the gallery (open-set scenario) cause false-matches with higher chances, which in turn degrades the overall biometric performance.

4.2.2.2 Workload Reduction: In order to evaluate the scalability of our hash look-up table, we compute in Tab. 7 the workload reduction (W) of our identification system according to the equation defined in Sect. 3.5 (see equation 3). Note some parameters such as θ and β are estimated in the Sect. 3.5. Penetration rate (*i.e.*, p) can be computed as is defined in the equation 4. Additionally, the execution time (*i.e.*, ψ) in seconds (or milliseconds) is reported for an entire identification transaction for each tested configuration of the proposed system.

In our experiments, a large number N of enrolled subjects is selected for this purpose. To that end, we join the FEI and FERET databases. In this case, one sample per subject is enrolled resulting in a total amount of N = 1,177 enrolled subjects. In addition, we calculate the workload W_B for the baseline (*i.e.*, $1 : N$ exhaustive search). For the baseline, ψ is approximately 14 minutes, p is assumed as 1.0, and β as 0.0, thereby resulting in a workload of $W_B = 88.27 \times 10^4$ representing 100%. Here, our system must satisfy the relation $0.0\% < W < 100\%$.

From Tab. 7, we can observe that there do exist a high dependence between the workload reduction (W) and the main parameters of the proposed system (*i.e.*, K and P): W decreases as K and P increase. In particular, our system is able to decrease W_B down to 0.09 for higher K and P values. Hence, a penetration rate of approximately 9×10^{-4} for $K \geq 256$ and $P \geq 2$ can be noted. In order to estimate W for a large number of enrolled subjects (*e.g.*, N = 1 million), we finally compute in Fig. 9, a linear regression for the worst (*i.e.*, $P = 1$) and best (*e.g.*, $P = 4$) cases over $K = 64$ and $K = 1024$ respectively, which reveals the linear relation between N and p. Therefore, the penetration rate p of our system in terms of percent (%) for a large dataset would be $p = 0.03$ for the best case and $p = 1.53$ for the worst case. Note that a low p leads to a low W.

4.2.2.3 Security analysis and privacy protection: We assume that our identification system could be attacked at three components of the system, i) the client device, ii) the communication channel between the client and the database, and iii) the database. More specifically, we follow a semi-honest security model where all the parties are constrained to follow the protocol and learn nothing beyond their own outputs, although, they might try to learn some information as possible. In particular, our scheme exploits the security provided by the BFV scheme based on the hardness of the Ring Learning [70], similar to Engelsma *et al.* [24]. Therefore, a ciphertext cannot be decrypted without getting access to the private key which is located only in the client entity. In addition, the hash codes which are utilised as



(a) LFW (O1) (b) LFW (O2)

Fig. 8. DET-curves of our best performing hash generation scheme (*i.e.*, AP) in an open-set scenario.

TABLE 7
Workload reduction results in terms of percent (%) for different parameter configurations over $N = 1,177$ enrolled subjects taken from FEI and FERET with $\theta = 750ms$ and $\beta = 0.003ms$. The workload and the execution time for the baseline system is $W_B = 88.27 \times 10^4$ and approximately 14 minutes respectively

K	Metrics	$P = 1$	$P = 2$	$P = 4$
64	γ	18.3906	1.3282	1.0020
	ψ	$\sim 14sec$	$\sim 750ms$	$\sim 750ms$
	p	156×10^{-4}	11×10^{-4}	8×10^{-4}
	W	1.56%	0.11%	0.09%
128	γ	9.1953	1.1444	1.0007
	ψ	$\sim 7sec$	$\sim 750ms$	$\sim 750ms$
	p	78×10^{-4}	10×10^{-4}	9×10^{-4}
	W	0.78%	0.1%	0.08%
256	γ	4.5977	1.0815	1.0022
	ψ	$\sim 3sec$	$\sim 750ms$	$\sim 750ms$
	p	39×10^{-4}	9×10^{-4}	9×10^{-4}
	W	0.39%	0.09%	0.09%
512	γ	2.2988	1.0591	1.0022
	ψ	$\sim 2sec$	$\sim 750ms$	$\sim 750ms$
	p	20×10^{-4}	9×10^{-4}	9×10^{-4}
	W	0.2%	0.09%	0.09%
1024	γ	1.1424	1.0148	1.0012
	ψ	$\sim 750ms$	$\sim 750ms$	$\sim 750ms$
	p	10×10^{-4}	9×10^{-4}	9×10^{-4}
	W	0.1%	0.09%	0.09%

keys in our hash look-up table, are computed as the binary representation of the codeword index and not directly from the biometric features. Hence, this does not leak information from biometric features of the subject. Further, in the case of supervision of an attacker on the data structure (*i.e.*, hash look-up table), it is relevant that the access to an entry in our hash look-up table is performed in $O(1)$. That is, it would become a hard task to get similarity information from the keys for any attacker since distance values are not revealed.

With respect to privacy protection, our system is able to fulfil the requirements from ISO/IEC 24745 [17]. In the

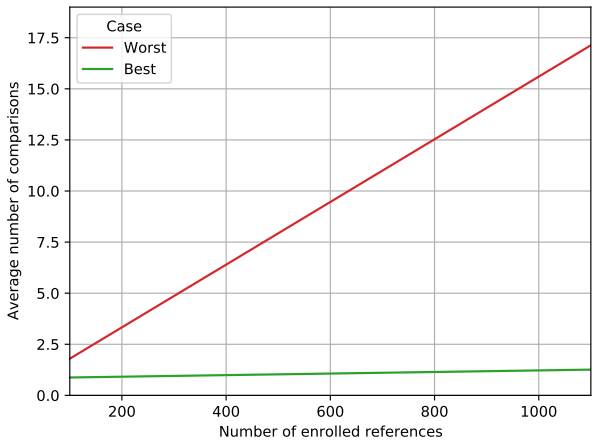


Fig. 9. Relation between the number of enrolled subjects and average number of comparisons.

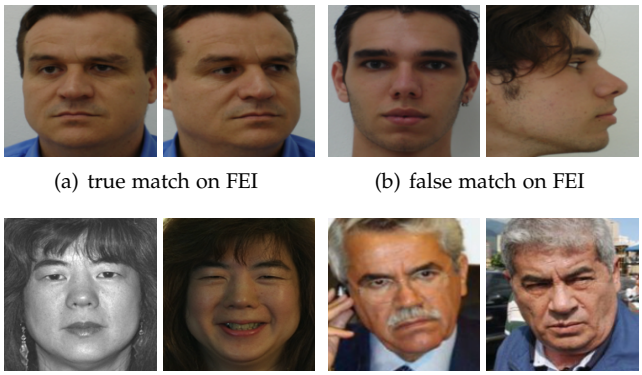
context of the keys, hash codes stored in the hash look-up table are not discriminative which in turn means that it is unlikely that they can be used to reconstruct original face templates. However, it should noted that if the hash codes are exposed, *i.e.*, information could be leaked. If two templates from a same subject result in a same generated hash code, cross-matching could potentially be performed. Also, an attacker could try to reconstruct a pre-image which may not necessarily identical to the original face. Hence, to further protect the hash codes in the hash look-up table, the proposed system could cryptographically hash the binary string to get a secure hash code by using *e.g.*, SHA256. This would prevent reconstruction attacks. As previously mentioned, unlinkabilty and renewability of hash codes can be achieved with MACs while additional key storage is required. Note that the application of such conventional cryptographic methods is only feasible since the proposed system enables an exact match of hash codes. Further, note that this extension to the proposed system does not have

any effects on its performance.

It is important to note that for suitable parameter configurations hash codes tend to be short. For example, for $K = 1,024$, and $P = 4$, the hash code length is $4\log(1024) = 40$ bits. Nevertheless, it is not realistic to achieve practical security from a cryptographic point of view, *i.e.*, more than 128 bits. This is because the entropy of a face embedding is considered to be much lower than what would be required for strong cryptographic protection. For instance, ArcFace extracts embeddings of 512 values. However, several researchers, *e.g.*, Gong *et al*. [71], have shown that the intrinsic dimensionality of those and other typical facial embeddings extracted by neural networks is much lower (in excess of an order of magnitude); *i.e.*, guessing the biometric template is a more feasible (although still very difficult) attack vector than guessing the encryption keys. This is further underlined by the high FPIRs at which face identification systems are operated at, *e.g.*, FPIR=0.1%. Consequently, the security and, hence, privacy protection, are also upper-bounded by attackers' effort of being falsely accepted by the system, the so-called false accept attack. Due to these reasons, the length of the extracted hash-codes is considered sufficient.

Focusing on the original face templates unlinkability can be guaranteed through FHE, where different set of keys for encryption are used. Irreversibility of the templates depends on the FHE parameters which support a high security level. Renewability is ensured, since new biometric templates can be obtained by changing the encryption keys. In the FHE scheme biometric performance can be maintained since comparisons in the encrypted domain are equal to those in the plaintext domain. It is also very important to note that the potential attack vectors of guessing the encryption keys and/or the biometric templates are by no means limited to homomorphic encryption in biometrics – other template security approaches including classic general-purpose encryption and dedicated biometric template protection schemes (*e.g.*, biometric cryptosystems or fuzzy vaults) likewise have to address those challenges.

4.2.2.4 Benchmark with state-of-the-art: A benchmark of the proposed method with the state-of-the-art on the LFW dataset is shown in Tab. 8. It can be observed that our system achieves a biometric performance which is comparable to the state-of-the-art. In particular, our system reports a FNIR100 = 2.97% and FNIR1000 = 34.99%, respectively, for $P = 2$ sub-spaces when images from known unknown subjects are used for the hash generation scheme training (*i.e.*, scenario $O1$). In contrast, a biometric performance degradation can be seen for a FPIR = 0.1% over $O2$: a FNIR1000 = 70.06% indicates the limitation of our scheme on the unseen data in challenging databases. In spite of this limitation, our identification system, unlike the proposal in [19], is capable of reducing the number of comparisons on the dataset, thereby achieving to a remarkable trade-off between privacy protection, biometric performance, and efficiency. It is important to highlight that the identification scheme described in [19] carried out an exhaustive search, which decreases its use for a real-time application. In addition, it should be noted, Random IoM and LIoM apply different strategies for training their compact hash codes while maintaining the protocol over the LFW database.



(a) true match on FEI (b) false match on FEI

(c) true match on FERET (d) false match on LFW

Fig. 10. Examples of (a) true match for slight variations in pose, (b) false match due to extreme pose variation, (c) true match for variations in color and expression, and (d) false match due to high similarity and low image quality.

4.2.2.5 Data exploration: In order to measure the robustness of the proposed system and to detect sources or errors we compute a clustering internal validation measure called Silhouette-coefficient [72]. This is done for the best performing hash generation scheme (*i.e.*, AP) on all used databases. The Silhouette-coefficient metric calculates the degree of separation (*i.e.*, inter-cluster) and cohesion (*i.e.*, intra-cluster) between clusters: scores close to 1 indicate a high separation between clusters while values close to -1 denote a high dispersion between the samples within a cluster and hence a high intra-class variation which could lead to an overlap between clusters. For the FEI, FERET, and LFW databases Silhouette-coefficients of 0.77, 0.82, and 0.24 are obtained.

Fig. 10, depicts cases of true and false matches. Firstly, it can be seen that the proposed system is robust to variations in facial expression, pose, or colour. However, extreme pose variations or low image quality can lead to false matches. We also may observe that the Silhouette-coefficient scores for those groups exhibiting less variation in pose are usually higher, *e.g.*, 0.9. Nevertheless, face images with extreme pose variations are expected to occur less frequently in cooperative scenarios. The highest Silhouette-coefficient score (*i.e.*, 0.82) is obtained on the FREET dataset, since it contains mostly high quality frontal-pose face images which is three times greater than the one reported for the LFW (*i.e.*, 0.24). The result achieved for the LFW confirms the impact of image quality over our hash generation scheme.

5 Conclusions

In this paper, we proposed a new privacy-preserving face identification system which allows indexing and retrieving candidate lists of protected face templates in $O(1)$. In particular, a hash generation scheme based on Product Quantisation was introduced in order to extract a stable hash code from facial images. Compact hash codes generated by said hash generation are used for efficient indexing via a hash look-up table. Corresponding face templates are encrypted through FHE and stored in the entries as protected reference templates. In the context of building a stable hash code, four different clustering algorithms were evaluated

TABLE 8
Benchmark, in terms of FPIR and FNIR (in %), of the best performing configuration of the proposed system with the state-of-the-art (without fusion scheme) in open-set scenarios on the LFW database.

System	WR Category	BTP Category	O1 (FNIR1000)	O1 (FNIR100)	O2 (FNIR1000)	O2 (FNIR100)
Random IoM [19]	Feature Transformation	Cancelable	40.47	**2.37**	**10.97**	2.37
LIoM [19]	Feature Transformation	Cancelable	45.81	2.43	14.37	**2.25**
Proposed ($P = 1$)	Pre-selection (hash look-up table)	FHE	36.40	2.68	74.07	2.60
Proposed ($P = 2$)	Pre-selection (hash look-up table)	FHE	**34.99**	2.97	70.06	2.80
Proposed ($P = 4$)	Pre-selection (hash look-up table)	FHE	37.63	3.76	88.22	3.10

(*i.e.*, K-means, K-medoids, GMM, and AP). Specifically, AP, which does not require to pre-define the number of clusters, achieved competitive results, *i.e.*, a remarkably low pre-selection error rate of 0.86%.

The experimental evaluation was carried out over different public datasets including the challenging LFW database. We evaluated the trade-off between privacy protection, security, biometric performance, and computational efficiency of the proposed system. In particular, we noted the capability of AP to maintain a high performance when the number of samples in their training varies. We showed the stability of AP by using only three samples for training. Experiments for several image quality ranges showed that face image quality has significant impact on the proposed system. However, the experimental evaluation over the LFW showed that for high-quality images (*i.e.*, FaceQNet scores greater than 0.5) reliable performance can be achieved. In addition, a benchmark with some state-of-the-art methods showed promising results of our system to reject known unknown subjects: a FNIR in the range of 2.68%-37.63% for a high-security threshold together with a remarkable workload reduction down to 0.1% was achieved, yielding reliable, secure, and fast identification system. Finally, this work indicates the need for introducing pre-selection-based WR strategies combined with FHE-based schemes to reduce workload and hence its feasibility in real applications.

In order to tackle some of the limitations of the proposed system, we plan to: i) improve the process of hash generation based on PQ with AP clustering. AP finds its best exemplars on a latent space of face embeddings (by using distance metrics). Keeping in mind that face embeddings in latent spaces could be distorted due to quality problems, probabilistic approaches could be implemented to analyse the dependence of the samples from a same subject and hence selecting the best exemplars in AP; ii) more stable hash generation may be achieved by combining PQ with compact binary representations and clustering in Hamming space, thereby alleviating the problem of the intra-class variability; iii) combine our pre-selection method (*i.e.*, hash look-up table) with other WR techniques such as feature transformation to improve the comparison time in the encrypted domain.

ACKNOWLEDGEMENTS

This work has in part received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 860813 - TReSPAsS-ETN and the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

REFERENCES

[1] European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice, "Eurodac storage capacity increased," https://www.eulisa.europa.eu/Newsroom/News/Pages/Eurodac-storage-capacity-increased.aspx, April 2016, last accessed: July 26, 2021.

[2] European Commission, "Smart borders," https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/smart-borders_en, 2018, last accessed: July 26, 2021.

[3] Gemalto, "DHS's automated biometric identification system IDENT - the heart of biometric visitor identification in the USA," https://www.gemalto.com/govt/customer-cases/ident-automated-biometric-identification-system, March 2019, last accessed: July 26, 2021.

[4] A. Dalwai, "Aadhaar technology and architecture: principles, design. best practices and key lessons," 2014.

[5] UIDAI, "Role of biometric technology in aadhaar enrollment," Unique Identification Authority of India, Tech. Rep., January 2012.

[6] Gemalto, "Automated Fingerprint Identification System (AFIS) - a short history," https://www.gemalto.com/govt/biometrics/afis-history, April 2019, last accessed: July 26, 2021.

[7] Federal Bureau of Investigation, "CODIS - NDIS statistics," https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics, June 2018, last accessed: July 26, 2021.

[8] S. Kundra, A. Dureja, and R. Bhatnagar, "The study of recent technologies used in e-passport system," in *2014 IEEE Global Humanitarian Technology Conf. South Asia Satellite (GHTC-SAS)*. IEEE, 2014, pp. 141–146.

[9] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 2382-37:2017 Information Technology - Vocabulary - Part 37: Biometrics*, International Organization for Standardization, 2017.

[10] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 19795-2:2006. Information Technology – Biometric Performance Testing and Reporting – Part 2: Testing methodologies for technology and scenario evaluation*, International Organization for Standardization and International Electrotechnical Committee, 2006.

[11] P. Drozdowski, C. Rathgeb, and C. Busch, "Computational workload in biometric identification systems: An overview," *IET Biometrics*, vol. 8, no. 6, pp. 351–368, November 2019.

[12] Unique Identification Authority of India, "Aadhaar dashboard," https://www.uidai.gov.in/aadhaar_dashboard/, 2021, last accessed: July 26, 2021.

[13] J. Daugman, "Biometric decision landscapes," University of Cambridge, Computer Laboratory, Tech. Rep., 2000.

[14] European Council, "Regulation of the european parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (general data protection regulation)," April 2016.

[15] M. Gomez-Barrero, J. Galbally, C. Rathgeb, and C. Busch, "General framework to evaluate unlinkability in biometric template protection systems," *IEEE Trans. on Information Forensics and Security*, vol. 13, no. 6, pp. 1406–1420, June 2018.

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

[17] ISO/IEC JTC1 SC27 Security Techniques, *ISO/IEC 24745:2011. Information Technology - Security Techniques - Biometric Information Protection*, International Organization for Standardization, 2011.

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

[19] X. Dong, S. Kim, Z. Jin, J. Hwang, S. Cho, and A. Teoh, "Open-set face identification with index-of-max hashing by learning," *Pattern Recognition*, vol. 103, p. 107277, 2020.

[20] T. Murakami, R. Fujita, T. Ohki, Y. Kaga, M. Fujio, and K. Takahashi, "Cancelable permutation-based indexing for secure and efficient biometric identification," *IEEE Access*, vol. 7, pp. 45 563–45 582, 2019.

[21] A. Sardar, S. Umer, C. Pero, and M. Nappi, "A novel cancelable facehashing technique based on non-invertible transformation with encryption and decryption template," *IEEE Access*, vol. 8, pp. 105 263–105 277, 2020.

[22] P. Drozdowski, N. Buchmann, C. Rathgeb, M. Margraf, and C. Busch, "On the application of homomorphic encryption to face identification," in *Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*, September 2019, pp. 1–8.

[23] M. Gomez-Barrero, J. Galbally, A. Morales, and J. Fierrez, "Privacy-preserving comparison of variable-length data with application to biometric template protection," *IEEE Access*, vol. 5, no. 1, pp. 8606–8619, December 2017.

[24] J. Engelsma, A. Jain, and V. Boddeti, "Hers: Homomorphically encrypted representation search," *arXiv preprint arXiv:2003.12197*, 2020.

[25] Y. Wang, J. Wan, J. Guo, Y.-M. Cheung, and P. Yuen, "Inference-based similarity search in randomized montgomery domains for privacy-preserving biometric identification," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 40, no. 7, pp. 1611–1624, 2017.

[26] H. Jegou, M. Douze, and C. Schmid, "Product quantization for nearest neighbor search," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 33, no. 1, pp. 117–128, 2010.

[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 and International Electrotechnical Committee, 2021.

[28] C. Thomaz and G. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image and vision computing*, vol. 28, no. 6, pp. 902–913, 2010.

[29] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[30] 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.

[31] A. Gionis, P. Indyk, R. Motwani *et al.*, "Similarity search in high dimensions via hashing," in *Vldb*, vol. 99, no. 6, 1999, pp. 518–529.

[32] A. Chaari, S. Lelandais, and M. Ahmed, "A pruning approach improving face identification systems," in *2009 Sixth IEEE Intl. Conf. on Advanced Video and Signal Based Surveillance*. IEEE, 2009, pp. 85–90.

[33] P. Mohanty, S. Sarkar, R. Kasturi, and P. Phillips, "Subspace approximation of face recognition algorithms: an empirical study," *IEEE Trans. on Information Forensics and Security*, vol. 3, no. 4, pp. 734–748, 2008.

[34] J. Gentile, N. Ratha, and J. Connell, "SLIC: Short-length iris codes," in *2009 IEEE 3rd Intl. Conf. on Biometrics: Theory, Applications, and Systems*. IEEE, 2009, pp. 1–5.

[35] M. Lim, A. Teoh, and J. Kim, "Biometric feature-type transformation: Making templates compatible for secret protection," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 77–87, 2015.

[36] N. Damer, P. Terhörst, A. Braun, and A. Kuijper, "Efficient, accurate, and rotation-invariant iris code," *IEEE Signal Processing Letters*, vol. 24, no. 8, pp. 1233–1237, 2017.

[37] Z. Jin, J. Hwang, Y.-L. Lai, S. Kim, and A. Teoh, "Ranking-based locality sensitive hashing-enabled cancelable biometrics: Index-of-max hashing," *IEEE Trans. on Information Forensics and Security*, vol. 13, no. 2, pp. 393–407, 2017.

[38] A. Gionis, P. Indyk, R. Motwani *et al.*, "Similarity search in high dimensions via hashing," in *Vldb*, vol. 99, no. 6, 1999, pp. 518–529.

[39] K. Cheung, A. Kong, D. Zhang, M. Kamel, J. You, and H.-W. Lam, "An analysis on accuracy of cancelable biometrics based on biohashing," in *Intl. Conf. on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, 2005, pp. 1168–1172.

[40] A. Lumini and L. Nanni, "An improved biohashing for human authentication," *Pattern Recognition*, vol. 40, no. 3, pp. 1057–1065, 2007.

[41] X. Dong, Z. Jin, A. Teoh, M. Tistarelli, and K. Wong, "On the reliability of cancelable biometrics: Revisit the irreversibility," *CoRR*, vol. abs/1910.07770, 2019. [Online]. Available: http://arxiv.org/abs/1910.07770

[42] H. Wang, X. Dong, Z. Jin, A. Teoh, and M. Tistarelli, "Interpretable security analysis of cancellable biometrics using constrained-optimized similarity-based attack," in *Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision*, 2020, pp. 70–77.

[43] S. Gong, V. Boddeti, and A. Jain, "On the intrinsic dimensionality of image representations," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2019, pp. 3987–3996.

[44] J. Fan and F. Vercauteren, "Somewhat practical fully homomorphic encryption." *IACR Cryptol. ePrint Arch.*, vol. 2012, p. 144, 2012.

[45] J. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in *Intl. Conf. on the Theory and Application of Cryptology and Information Security*. Springer, 2017, pp. 409–437.

[46] C. Gentry and S. Halevi, "Implementing gentry's fully-homomorphic encryption scheme," in *Annual Intl. Conf. on the theory and applications of cryptographic techniques*. Springer, 2011, pp. 129–148.

[47] O. Goldreich, *Foundations of cryptography: volume 2, basic applications*. Cambridge university press, 2009.

[48] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Proc. Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[49] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proc. of the IEEE Conf. on computer vision and pattern recognition*, 2015, pp. 815–823.

[50] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2019, pp. 4690–4699.

[51] 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 Intl. Conf. on automatic face & gesture recognition (FG 2018)*. IEEE, 2018, pp. 67–74.

[52] D. Wang, C. Otto, and A. Jain, "Face search at scale," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1122–1136, 2016.

[53] C. Otto, D. Wang, and A. Jain, "Clustering millions of faces by identity," *IEEE trans. on pattern analysis and machine intelligence*, vol. 40, no. 2, pp. 289–303, 2017.

[54] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proc. of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA, 1967, pp. 281–297.

[55] L. Kaufman and P. Rousseeuw, "Clustering by means of medoids in statistical data analysis based on the l1–norm and related methods.(y. dodge, dü.) reports of the faculty of mathematics and informatics," *Delft University of Technology*, 1987.

[56] N. Shental, A. Bar-Hillel, T. Hertz, and D. Weinshall, "Computing gaussian mixture models with em using equivalence constraints," *Advances in neural information processing systems*, vol. 16, no. 8, pp. 465–472, 2004.

[57] B. Frey and D. Dueck, "Clustering by passing messages between data points," *science*, vol. 315, no. 5814, pp. 972–976, 2007.

[58] Y. Cheung, "k-means: A new generalized k-means clustering algorithm," *Pattern Recognition Letters*, vol. 24, no. 15, pp. 2883–2893, 2003.

[59] K. He, F. Wen, and J. Sun, "K-means hashing: An affinity-preserving quantization method for learning binary compact codes," in *Proc. Intl. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 2938–2945.

[60] L.-J. González-Soler, M. Gomez-Barrero, L. Chang, A. Pérez-Suárez, and C. Busch, "Fingerprint presentation attack detection

based on local features encoding for unknown attacks," *IEEE Access*, 2021.
[61] Z. Brakerski, C. Gentry, and S. Halevi, "Packed ciphertexts in lwe-based homomorphic encryption," in *International Workshop on Public Key Cryptography*. Springer, 2013, pp. 1–13.
[62] N. Smart and F. Vercauteren, "Fully homomorphic simd operations," *Designs, codes and cryptography*, vol. 71, no. 1, pp. 57–81, 2014.
[63] J. Cheon, H. Chung, M. Kim, and K.-W. Lee, "Ghostshell: Secure biometric authentication using integrity-based homomorphic evaluations." *IACR Cryptol. ePrint Arch.*, vol. 2016, p. 484, 2016.
[64] P. Drozdowski, F. Struck, C. Rathgeb, and C. Busch, "Benchmarking binarisation schemes for deep face templates," in *Intl. Conf. on Image Processing (ICIP)*. IEEE, October 2018, pp. 1–5.
[65] "Microsoft seal (release 3.2)," https://github.com/Microsoft/SEAL, February 2019, microsoft Research, Redmond, WA.
[66] J. Kolberg, P. Drozdowski, M. Gomez-Barrero, C. Rathgeb, and C. Busch, "Efficiency analysis of post-quantum-secure face template protection schemes based on homomorphic encryption," in *Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*, 2020.
[67] P. Drozdowski, C. Rathgeb, B.-A. Mokroß, and C. Busch, "Multi-biometric identification with cascading database filtering," *Trans. on Biometrics, Behavior, and Identity Science (TBIOM)*, vol. 2, no. 3, pp. 210–222, July 2020.
[68] J. Hernandez-Ortega, J. Galbally, J. Fierrez, R. Haraksim, and L. Beslay, "Faceqnet: Quality assessment for face recognition based on deep learning," in *2019 Intl. Conf. on Biometrics (ICB)*. IEEE, 2019, pp. 1–8.
[69] Y. Shi and A. Jain, "Probabilistic face embeddings," in *Proc. of the IEEE/CVF Intl. Conf. on Computer Vision*, 2019, pp. 6902–6911.
[70] V. Lyubashevsky, C. Peikert, and O. Regev, "On ideal lattices and learning with errors over rings," in *Annual Intl. Conf. on the Theory and Applications of Cryptographic Techniques*. Springer, 2010, pp. 1–23.
[71] S. Gong, V. N. Boddeti, and A. K. Jain, "On the intrinsic dimensionality of image representations," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2019, pp. 3987–3996.
[72] P. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

B.Sc. Dailé Osorio received the B.Sc. degree in Computer Science from the Technological University of Havana, in 2014. She joined the Advanced Technologies Application Center (CENATAV), Havana, Cuba, for computer science graduate training. She is currently pursuing the Ph.D. degree at the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. She is a member of the da/sec – Biometrics and Internet Security Research Group and the National Research Center for Applied Cybersecurity (ATHENE), Germany. Her principal research interests are focused in areas of Pattern Recognition, Biometrics and Machine Learning, specifically, biometric indexing and privacy-enhancing technologies.

Dr. Christian Rathgeb is a Senior Researcher with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. He is a Principal Investigator in the National Research Center for Applied Cybersecurity ATHENE. His research includes pattern recognition, iris and face recognition, security aspects of biometric systems, secure process design and privacy enhancing technologies for biometric systems. He co-authored over 100 technical papers in the field of biometrics. He is a winner of the EAB - European Biometrics Research Award 2012, the Austrian Award of Excellence 2012, Best Poster Paper Awards (IJCB'11, IJCB'14, ICB'15) and the Best Paper Award Bronze (ICB'18). He is a member of the European Association for Biometrics (EAB), a Program Chair of the International Conference of the Biometrics Special Interest Group (BIOSIG) and an editorial board member of IET Biometrics (IET BMT). He has served for various program committees and conferences (*e.g.*ICB, IJCB, BIOSIG, IWBF) and journals as a reviewer (*e.g.*IEEE TIFS, IEEE TBIOM, IET BMT).

Dr. Pawel Drozdowski is a Senior Researcher with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. His research interests include biometrics, information security and privacy, pattern recognition, and algorithmic fairness. He co-authored over 25 technical publications in the field of biometrics. He won the Best Student Paper Runner-Up Award (WIFS'18) and Best Poster Award (BIOSIG'19). He is a member of the European Association for Biometrics (EAB) and represents the German Institute for Standardization (DIN) in ISO/IEC SC37 JTC1 SC37 on biometrics.

Prof. Dr. Christoph Busch is member of the Norwegian University of Science and Technology (NTNU), Norway. He holds a joint appointment with Hochschule Darmstadt (HDA), Germany. Further he lectures Biometric Systems at Denmark's DTU since 2007. On behalf of the German BSI he has been the coordinator for the project series BioIS, BioFace, BioFinger, BioKeyS Pilot-DB, KBEinweg and NFIQ2.0. He was/is partner of the EU projects 3D-Face, FIDELITY, TURBINE, SOTAMD, RESPECT, TReSPsS, iMARS and others. He is also principal investigator in the German National Research Center for Applied Cybersecurity (ATHENE) and is co-founder of the European Association for Biometrics (EAB). Christoph co-authored more than 500 technical papers and has been a speaker at international conferences. He is member of the editorial board of the IET journal on Biometrics and formerly of the IEEE TIFS journal. Furthermore, he chairs the TeleTrusT biometrics working group as well as the German standardization body on Biometrics and is convenor of WG3 in ISO/IEC JTC1 SC37.