Information Sciences 370–371 (2016) 18–32

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

# Unlinkable and irreversible biometric template protection based on bloom filters

Marta Gomez-Barrero[a,*], Christian Rathgeb[b], Javier Galbally[c], Christoph Busch[b,d], Julian Fierrez[a]

[a] ATVS - Biometric Recognition Group, EPS, Universidad Autonoma de Madrid, Spain
[b] da/sec - Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany
[c] Inst. for the Protection and Security of the Citizen, European Commission - JRC, Italy
[d] NISlab, Norwegian University of Science and Technology, NTNU, Gjøvik, Norway

ARTICLE INFO

*Article history:*
Received 9 March 2016
Revised 28 May 2016
Accepted 26 June 2016
Available online 29 June 2016

*Keywords:*
Unlinkability
Template protection
Privacy
Biometrics
Face

ABSTRACT

Deployments of biometric technologies are already widely disseminated in numerous large-scale nation-wide projects. Since the protection of biometric reference data is of particular concern in order to safeguard individuals' privacy, biometric template protection schemes are designed to handle biometric reference data in an irreversible and unlinkable manner. In past years, schemes based on Bloom filters have been introduced and applied to various characteristics. However, thorough security analyses have exposed the original concept to be vulnerable to cross-matching attacks.

In this article we present a general framework for the evaluation of unlinkability in biometric template protection schemes, as well as an improved, unlinkable and irreversible, system based on Bloom filters. In order to generate cross-matching resistant protected templates we re-design the original scheme and propose an additional, easily integrable, processing step, which is referred to as structure-preserving feature re-arrangement. The improved system is thoroughly evaluated on the publicly available face corpus of the BioSecure Multimodal Database. It is shown that the proposed scheme maintains the biometric performance of the unprotected system. Moreover, cross-matching resistance is achieved in the presence of existing attacks, considering adversary models where potential attackers are in possession of protected biometric templates as well as secret credentials.

## 1. Introduction

Biometrics refers to automated recognition of individuals based on their behavioral and biological characteristics, e.g. fingerprint or face [18]. Providing a strong link between an identity and its owner, it is generally conceded that a substitute to biometrics for positive identification in integrated security applications is non-existent. However, unprotected storage of biometric reference data (templates) poses serious privacy threats, e.g. identity theft, cross-matching, or limited renewability. Moreover, biometric data is considered sensitive data, as defined in European Union (EU) data protection directive IP/12/46[1] [11], which means that the use biometric data is subjected to right of privacy preservation. *Biometric template*

* Corresponding author.
*E-mail addresses:* marta.barrero@uam.es (M. Gomez-Barrero), christian.rathgeb@h-da.de (C. Rathgeb), javier.galbally@jrc.ec.europa.eu (J. Galbally), christoph.busch@ntnu.no (C. Busch), julian.fierrez@uam.es (J. Fierrez).

[1] http://ec.europa.eu/justice/data-protection/reform/index_en.htm

http://dx.doi.org/10.1016/j.ins.2016.06.046
0020-0255/© 2016 Elsevier Inc. All rights reserved.

*protection* technologies [8,26,29,31,35], which are commonly categorized as *biometric cryptosystems* and *cancelable biometrics*, offer solutions to privacy preserving biometric authentication. Cancelable biometrics consist of intentional, repeatable distortions of biometric signals based on transforms that provide a comparison of biometric templates in the transformed domain, i.e. biometric templates are permanently protected [32]. Given a biometric datum **M** (i.e., raw biometric signal), a pre-chosen (application-specific) *secret key* **T** is incorporated as a parameter of a *Pseudonymous Identifier Encoder* (*PIE*) in order to generate the corresponding cancelable template $\mathbf{C} = PIE(\mathbf{M}, \mathbf{T})$. A corresponding *Pseudonymous Identifier Comparator* (*PIC*) is employed to compare pairs of cancelable templates at the time of authentication. In accordance with the ISO/IEC IS 24745 [16], cancelable biometrics are required to meet the two major requirements of *irreversibility* and *unlinkability*:

1. *Irreversibility*: knowledge of a cancelable template **C** and corresponding key **T** can not be exploited to reconstruct a biometric signal $\mathbf{M}'$ which positively matches the original biometric sample **M**. This property prevents the abuse of stored biometric data for launching spoof or replay attacks, thereby improving the security of biometric systems [25].
2. *Unlinkability*: given **M**, it must be feasible to generate different versions of cancelable templates $\mathbf{C}^1, \mathbf{C}^2, \dots, \mathbf{C}^U$ by incorporating different keys $\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_U$, $\mathbf{C}^i = PIE(\mathbf{M}, \mathbf{T}_i)$, $i = 1, \dots, U$, so that those templates cannot be linked to a single subject. This property guarantees the privacy of a subject when he is registered in different applications with the same biometric trait, preventing *cross-matching attacks*, and also allows issuing new credentials in case a protected template is stolen.

In addition to the irreversibility and unlinkability properties, biometric template protection approaches should not affect other important performance factors of conventional biometric recognition systems [36]. For instance, accuracy of unprotected systems should be preserved and authentication speed should be comparable in order to enable fast identification.

While numerous biometric template protection schemes have been introduced [29,35], in most cases only biometric performance and irreversibility are analysed. Unlinkability, on the other hand, is most frequently only partially studied. This has led to a lack of an appropriate evaluation framework for this key property of biometric template protection schemes. In this article, we propose a new framework for the evaluation of the unlinkability provided by protected templates.

Then, we will present and evaluate a new unlinkable and irreversible biometric template protection system. Among the different schemes proposed, the present work will focus on the use of Bloom Filters for the protection of biometric templates and, in particular, on their application to the field of cancelable face biometrics. Biometric template protection based on Bloom filters was introduced in [33]. In contrast to other template protection approaches, it is not specific for a single characteristic since it has been successfully applied to iris [33], face [12] or fingerprint [1,24]. The scheme is designed to map biometric features to an irreversible representation, i.e. Bloom filters. Experimental evaluations have shown that the proposed scheme is capable of maintaining biometric performance and fast comparison of compact protected templates. Moreover, the concept can be utilized for multi-biometric template protection, where fusion is performed at feature level [34].

In the original Bloom filter template protection approach [33], in order to achieve unlinkability, the authors suggested to incorporate rather short secret keys, e.g. $|\mathbf{T}| \leq 2^{10}$ in [33], which further transforms parts of biometric features in a linear manner. Recently, a security analysis of the entire concept, in particular of the original iris-based system [33], was presented in [15]. While the irreversibility property of the system has been confirmed, it is shown that the initial scheme is vulnerable to cross-matching attacks. In [7] it has been demonstrated that suggested key-spaces are of insufficient size enabling brute-force cross-matching attacks. In addition, it has been shown that the irreversibility property of Bloom filter-based transforms depends on the nature of biometric data.

Taking into account the aforementioned issues, the main contributions of this article can be summarised as follows:

- A new framework for the systematic analysis of the unlinkability of biometric template protection schemes. Due to the lack of an appropriate metric for the unlinkability of the templates, in most related works this property has not been properly analysed. The development of a new framework for the analysis of the unlinkability of the templates, including two new measures to study both the unlinkability for each particular score and for a system as a whole, has been developed and applied to the proposed template protection scheme.
- An improved unlinkable and irreversible Bloom filter-based template protection scheme. Building upon the original concept of Bloom filter-based template protection proposed in [33], which provided irreversible templates, we introduce a *Structure-Preserving Feature Re-Arrangement* to produce irreversible and unlinkable templates in accordance with the ISO/IEC International Standard 24745 [16] on biometric information protection.

Experimental evaluations are carried out on the face corpus of the publicly available BioSecure Multimodal Database [27], using the free signal and image processing toolbox Bob [4], in order to generate fully reproducible research. The improved Bloom filter face biometric template protection scheme based on the original system introduced in [12] is thoroughly analysed with respect to irreversibility and unlinkability. Moreover, robustness to proposed attacks is verified, showing that fully unlinkable protected templates are achieved, considering two different adversary models, i.e. *advanced model* and *full disclosure model*. While in the former model, the attacker has full knowledge of the applied template protection algorithm and has access to protected templates, in the latter model an eventual attacker is in possession of the corresponding secret keys, too. In addition, biometric performance (accuracy) is maintained compared to the original unprotected face recognition system, confirming the soundness of the proposed approaches.

The remainder of this article is organized as follows: Section 2 summarizes related works with respect to cancelable face biometrics and Bloom filter-based template protection. Section 3 introduces the new unlikability analysis framework. A detailed description of the improved system is given in Section 4. In Section 5, potential attacks to the original Bloom filter-based scheme are described. The applied experimental protocol is summarized in Section 6. Experimental evaluations are presented in Section 7. Finally, conclusions are drawn in Section 8.

## 2. Related works

### 2.1. Cancelable face biometrics

Ratha et al. [32] were the first to introduce the concept of cancelable biometrics applying non-invertible transforms in the image domain. At enrolment, a non-invertible transform (e.g. surface folding) is applied to a facial image using application-dependent parameters. During authentication, probe images are transformed employing the same parameters and compared to the stored reference. In [6], cryptographically secure biotokens are proposed and applied to existing face recognition schemes, such as Principal Component Analysis (PCA). The key idea is to split biometric features into a stable part and an unstable part. Subsequently, stable parts are encrypted and unstable parts are obscured applying non-invertible projections. In the vast majority of approaches to cancelable biometrics, revocability is provided by incorporating secret credentials, e.g. random numbers. Consequently, security evaluations have to be performed under the "stolen-secret scenario", where and impostor is in possession of valid secrets.

In [38] a technique applied to face biometrics called "BioHashing" was introduced. Basically, the BioHashing approach operates as a key-binding scheme, using secret subject-specific tokens (unlike public auxiliary data) at authentication. Prior to the key-binding step, secret tokens are blended with biometric data to derive a distorted biometric template, i.e., BioHashing represents an instance of "Biometric Salting" [35]. In most biometric salting approaches, subject-specific keys are incorporated while experiments are performed under the non-stolen-secret scenario omitting the actual biometric performance of the system. In the field of biometric security, the *stolen-token scenario* refers to the case when a genuine subject's token is "stolen" and utilized by an imposter to perform zero-effort false-accept attempts, i.e. prior verification biometric features extracted from a potential attacker are transformed with secrets of the account they want to gain access to. In contrast the *non-stolen-token scenario* refers to the case where impostor templates are protected with randomly generated keys prior to comparison, which artificially improves the biometric performance of the system. It is important to note that, if at all, existing works consider the stolen-token scenario mostly for performance evaluation, and not for attacks on irreversibility or unlinkability. In a follow-up publication [37], a significant degradation of biometric performance is reported for the stolen-token scenario. In [21] subject-specific random projections are applied to PCA-based face features followed by an error minimizing template transform. Again, performance evaluations under the stolen-token scenario are omitted.

More recently, in [28], two face images are mixed in order to protect the subject's privacy hiding the gender information, while retaining their discriminative power for verification. In [9], a double sum procedure is carried over the attributes to provide cancelable face and voice templates. In [19], a simplified shielding function is applied to rotation invariant neighbour-based local binary pattern features (RINLBP) extracted from face images. In [30], feature level fusion of different facial features is applied to generate the cancelable template, based on the multi-fold random projection and fuzzy communication scheme.

Other recent template protection approaches for different characteristics include [2,3,23,39].

Finally, it is important to note, that a fair comparison between the afore mentioned schemes is hard to establish. Each system was evaluated on different, and mostly small, databases, under different scenarios. Moreover, approaches show different requirements, such as multi-instance enrolment or mandatory pre-alignment of facial images. Finally, in most cases, even if no attacks have yet been proposed, no thorough irreversibility and/or unlinkability analysis has been performed.

### 2.2. Bloom filter-based cancelable biometrics

The concept of Bloom filter-based template protection was introduced by Rathgeb et al. [33] in order to achieve cancelable iris biometrics. Generic iris recognition systems extract binary feature vectors based on a row-wise analysis of normalized iris textures, i.e. iris-codes typically represent two-dimensional binary feature vectors. It is proposed in [33] to divide the two-dimensional binary feature matrix into $n\!-\!blocks$ blocks of equal size, where each block consists of $n\!-\!bits \times n\!-\!words$ bits. A Bloom filter [5] is represented as a binary array $\mathbf{b}$ of length $2^{n-bits}$, where initially all bits are set to zero. From each block, a Bloom filter is extracted such that the transformed iris-code $\mathbf{C}$ consists of $n\!-\!blocks$ separate Bloom filters, $\mathbf{C} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{n-blocks}\}$. In order to map one block to a Bloom filter, the entire sequence of columns of each block, which are referred to as *words*, is successively transformed to their decimal indexes which are set to one in the corresponding Bloom filter. Each bit of a Bloom filter can be set to one multiple times, but only the first change has an effect. In order to achieve unlinkability, it is suggested to XOR each word $\mathbf{w}_i$, $i = 1, \ldots, n\!-\!words$, with a single application-specific key $\mathbf{T} \in [0, \ldots, 2^{n-bits} - 1]$.

As Bloom filter indexes are visible to an attacker, the reconstruction of the corresponding binary block involves an arrangement of $|\mathbf{b}| \leq n\!-\!words$ different words to a binary block of length $n\!-\!words$, where $|\mathbf{b}|$ represents the number of activated indexes of the Bloom filter $\mathbf{b}$. By the inclusion-exclusion principle, the total number of possible sequences $n\!-\!seq$

resulting in the same binary feature block is estimated as,

$$n\text{−}seq = \sum_{i=1}^{|\mathbf{b}|} (-1)^{|\mathbf{b}|-i} \binom{|\mathbf{b}|}{i} i^{n\text{−}words}. \tag{1}$$

Even small values of $|\mathbf{b}|$ yield relatively large values of $n\text{−}seq$, i.e. irreversibility is achieved [15].

The comparison between two Bloom filter-based templates $\mathbf{C}$ and $\mathbf{C}'$ is defined as the sum of all pairwise comparisons of corresponding Bloom filters, $\mathbf{b}_i$, $\mathbf{b}'_i$, $i = 1, \ldots, n\text{−}blocks$. Since Bloom filters comprise a variable number of ones (depending on the number of identical words within processed blocks), as proposed in [33], the dissimilarity between two protected templates can be efficiently estimated as,

$$s = PIC(\mathbf{C}^1, \mathbf{C}^2) = \frac{1}{n\text{−}blocks} \sum_{i=1}^{n\text{−}blocks} DS(\mathbf{b}_i^1, \mathbf{b}_i^2) = \frac{|\mathbf{b}_i^1 \oplus \mathbf{b}_i^2|}{|\mathbf{b}_i^1| + |\mathbf{b}_i^2|} \tag{2}$$

where the XOR operator counts the number of disagreeing bits, which is normalized by the Hamming weight (*HW*) of both Bloom filters $\mathbf{b}_i^1, \mathbf{b}_i^2$.

In [33] this concept has been applied to iris-codes maintaining recognition performance. In [12] protected facial templates are generated based on the above concept. Again, biometric performance is preserved. Moreover, it has been shown that the concept can be utilized to achieve protected biometric fusion [34], which further improves biometric performance as well as privacy protection. In addition, it has been shown that the proposed *PIC*, which represents a Hamming distance (*HD*) based comparator, enables an efficient biometric identification.

## 3. New framework for unlinkability analysis

In order to provide unlinkability as defined in Section 1, secret keys are commonly introduced into template protection schemes. The *key space size* $|\mathbf{T}|$ is thus required to be large enough such that brute force attacks on the key space should at least be as hard as a false acceptance attack, i.e. $|\mathbf{T}| \geq FMR^{-1}$ [10], where *FMR* is the False Match Rate of the protected system. As a consequence, in order to utilize the entire space of secret keys, a small distance between two keys should cause a large distance between the resulting protected templates.

An additional threat can arise from *linkage* or *cross-matching attacks*, where an eventual attacker is in possession of two protected templates $\mathbf{C}^1 = PIE(\mathbf{M}_1, \mathbf{T}_1)$, and $\mathbf{C}^2 = PIE(\mathbf{M}_2, \mathbf{T}_2)$, with $\mathbf{T}_1 \neq \mathbf{T}_2$. His goal is to determine whether both protected templates, $\mathbf{C}^1$ and $\mathbf{C}^2$, conceal the same biometric datum $\mathbf{M}$ (or different samples of biometric data extracted from the same biometric instance – e.g., the same left index finger).

To prevent such attacks, the dissimilarity score between those templates is required to be higher than a certain decision threshold $\tau$, used to take a final non-match verification decision: $s = PIC(\mathbf{C}^1, \mathbf{C}^2) > \tau$. Furthermore, given two biometric samples $\mathbf{M}_1$ and $\mathbf{M}_2$ obtained from different biometric instances, and two different keys $\mathbf{T}_1$ and $\mathbf{T}_2$, the following equations to compute the dissimilarity score $s$ should hold:

$$s = PIC(\mathbf{C}_1, \mathbf{C}_2) > \tau \begin{cases} \mathbf{C}^1 = PIE(\mathbf{M}_1, \mathbf{T}_1), \mathbf{C}^2 = PIE(\mathbf{M}_1, \mathbf{T}_2), \\ \mathbf{C}^1 = PIE(\mathbf{M}_1, \mathbf{T}_1), \mathbf{C}^2 = PIE(\mathbf{M}_2, \mathbf{T}_1). \end{cases} \tag{3}$$

As we will explain below, in order for Eq. 3 to hold, there has to be a specific overlap between the inter-class distributions of non-mated comparisons using different keys and the score distribution obtained by comparing identical biometric instances protected with different keys.

To extend formality to the problem being addressed, some mathematical notations are introduced in this section. Let us define the following hypothesis:

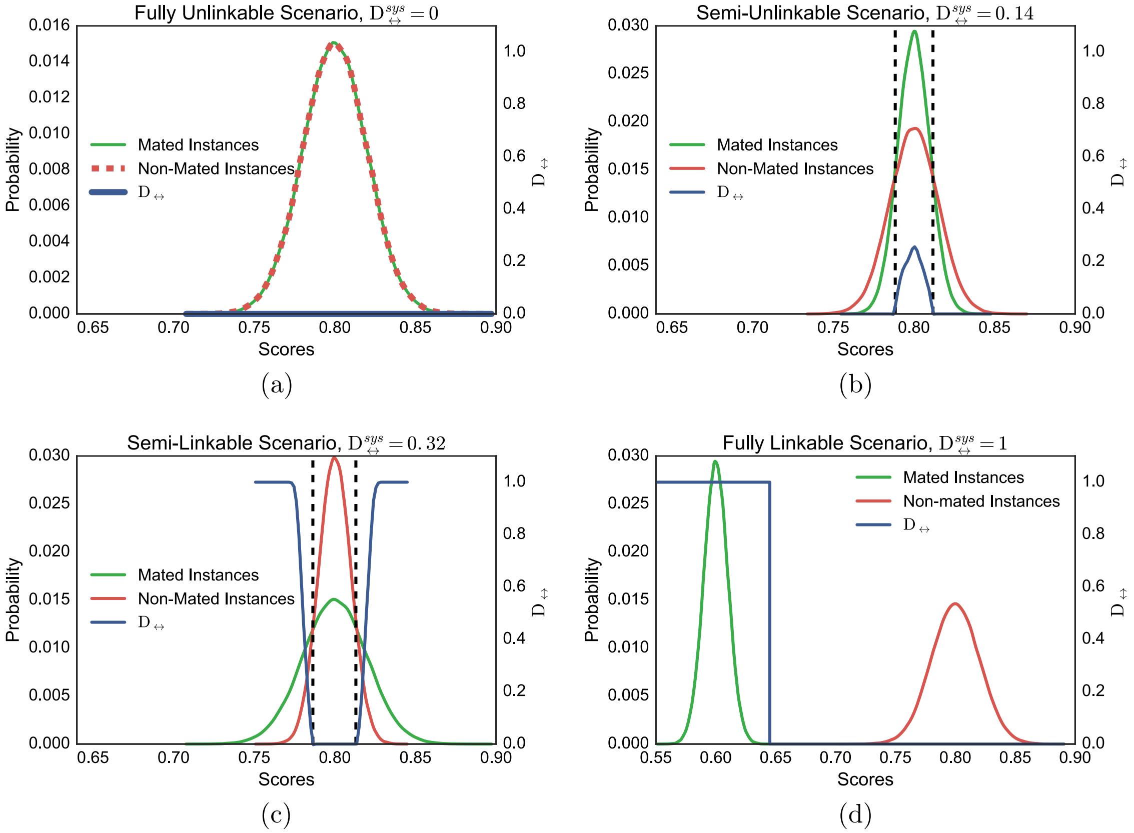
$$H_m = \{\text{both templates belong to mated instances}\} \tag{4}$$

$$H_{nm} = \{\text{both templates belong to non-mated instances}\} \tag{5}$$

Two types of score distributions will be analysed for the assessment of the unlinkability provided by protected templates:

- *Mated instances*: scores computed from templates extracted from different samples of a single instance of the same subject using different keys. It represents the probabilities $p(s|H_m)$, where $s$ is the dissimilarity score between two templates.
- *Non-mated instances*: scores yielded by templates generated from samples of different instances using different keys. It represents $p(s|H_{nm})$.

In this context, we assume that the attacker: *i*) is in possession of two protected templates $\mathbf{C}^1 = PIE(\mathbf{M}_1, \mathbf{T}_1)$, and $\mathbf{C}^2 = PIE(\mathbf{M}_2, \mathbf{T}_2)$, where $\mathbf{T}_1 \neq \mathbf{T}_2$, *ii*) can access the similarity score between them, $s = PIC(\mathbf{C}^1, \mathbf{C}^2)$, and *iii*) knows the *Mated instances* and *Non-mated instances* distributions.



(a)

(b)

(c)

(d)

**Fig. 1.** Examples of *mated instances* (green) and *non-mated instances* (red) distributions yielded by (a) fully unlinkable, (b) semi-unlinkable, (c) semi-linkable, and (d) fully linkable templates. While the blue curve represents the proposed unlinkability measure $\mathrm{D}_{\leftrightarrow}(s)$ for each possible score value, $\mathrm{D}_{\leftrightarrow}^{sys}$ gives an estimation of the unlinkability level of the whole system independently of the score range. The dashed black line represents $LR(s) = 1$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Traditionally, in order to compare the aforementioned distributions, the difference between probability densities has been estimated in terms of the Kullback–Leibler ($KL$) divergence [22] between two discrete distributions, $P$ and $Q$, which is defined as:

$$D_{KL}(P||Q) = \sum_{s} P(s) \ln\left(\frac{P(s)}{Q(s)}\right) \tag{6}$$

where $D_{KL} \geq 0$, and $D_{KL} = 0$ holds iff $P \simeq Q$, i.e. the smaller $D_{KL}$, the higher the similarity between distributions.

However, this measure is not appropriate due to three main reasons: *i*) it gives only an overall measure of the unlinkability of the system, not being possible to measure the level of unlinkability for different ranges of the similarity scores, *ii*) it is not bounded, thus making it difficult to compare the unlinkability of different systems, and *iii*) it is not defined for $Q(s) = 0$ if $P(s) \neq 0$, hence not taking into account important ranges of scores, or not being at all defined for fully separable distributions.

As a consequence, we need a new framework to evaluate the degree of unlinkability of such scenarios. To that end, we propose two different measures: $\mathrm{D}_{\leftrightarrow}^{sys}$ and $\mathrm{D}_{\leftrightarrow}(s)$. On the one hand, $\mathrm{D}_{\leftrightarrow}^{sys} \in [0, 1]$ gives an estimation of the linkability of a system as a whole, *independently* of the score. Accordingly, this metric is appropriate for example to compare the unlinkability level of two systems as a whole. This way, if a system has $\mathrm{D}_{\leftrightarrow}^{sys} = 1$ (i.e., case in which both the *Mated instances* and *Non-mated instances* distributions have no overlap, as shown in Fig. 1d), it means that it is fully linkable in all its score range. That is, if a cross-matching attack is carried out on the system between two protected templates $\mathbf{C}^1$ and $\mathbf{C}^2$, independently of the score produced, the attacker can know (with almost all certainty) if they conceal or not to same instance. Similarly, $\mathrm{D}_{\leftrightarrow}^{sys} = 0$ (i.e., Fig. 1a, where both score distributions totally overlap) means that the system is fully unlinkable for the whole score range. That is, independently of the score produced in a cross-matching attack, it is equally probable that the two templates come from the same instance ($H_m$) than from different instances ($H_{nm}$). All intermediate values of $\mathrm{D}_{\leftrightarrow}^{sys}$ between 0 and 1 report a decreasing degree of unlinkability (i.e., increasing degree of linkability).

On the other hand, $\mathrm{D}_{\leftrightarrow}(s) \in [0, 1]$ gives an estimation of the linkability of a system for a *specific score*. As such, this metric is appropriate to analyse within one system in which parts of the score range it fails to provide unlinkability. This way, if for a specific score $s_0$, a system yields $\mathrm{D}_{\leftrightarrow}(s_0) = 1$, it means that, *in case* a cross-matching attack produced $s_0$, the attacker would be able to link both templates $\mathbf{C}^1$ and $\mathbf{C}^2$ to the same user with almost all certainty. On the other hand, $\mathrm{D}_{\leftrightarrow}(s_0) = 0$ should be interpreted as full unlinkability for that particular score. In other words, *if* $s_0$ were produced in a cross-matching attack, the probability that both templates came from the same instance or from different instances would be the same. All intermediate values of $\mathrm{D}_{\leftrightarrow}(s)$ between 0 and 1 report a decreasing degree of unlinkability (i.e., increasing degree of linkability).

It should be noted that both measures yield values in a closed range, in opposition to $D_{KL}$, in order to allow a more straightforward comparison of different schemes. Next, we describe how both metrics, $\mathrm{D}_{\leftrightarrow}(s)$ and $\mathrm{D}_{\leftrightarrow}^{sys}$, are computed. Furthermore, to illustrate the different levels of unlinkability that templates can achieve, four different scenarios, which are described in the following, are shown in Fig. 1, where the *Mated instances* distribution is depicted in green and the *Non-mated instances* distribution in red, and the newly proposed $\mathrm{D}_{\leftrightarrow}(s)$ in blue. A fully unlinkable scenario is shown in Fig. 1a, where both distributions are identical. In this case, no decision can be made on whether, for a given score, the templates protect the same identity.

A semi-unlinkable scenario is shown in Fig. 1b, where the *Mated instances* distribution is enclosed within the *Non-mated instances* curve. As we may observe, for score values in [0.79, 0.81] we can state with some certainty that both templates are more likely to belong to the same instance. On the other hand, if the score is out of that range, the attacker can assume that such templates belong to different instances with a higher probability. Similarly, if he were able to compare a protected template with several references enrolled in the system in order to find the template concealing the same identity, he could discard templates yielding scores out of the aforementioned range and hence reduce the domain of his search.

A semi-linkable scenario is shown in Fig. 1c, where the *Mated instances* distribution spans further than the *Non-mated instances* curve. More specifically, if the score is out of the range [0.79, 0.81], the probability of both templates belonging to different instances is almost zero. As a consequence, we can assume with almost all certainty that both templates protect the same instance, thus making the templates linkable.

A fully-linkable scenario is shown in Fig. 1d, where the *Mated instances* and *Non-mated instances* distributions are fully separable. Therefore, the attacker can make a decision with almost all certainty for all scores.

### 3.1. Computation of $\mathrm{D}_{\leftrightarrow}(s)$ and $\mathrm{D}_{\leftrightarrow}^{sys}$

Inspired in the analysis of biometric forensic evidence [13], likelihood ratios can be used to give an estimation of those certainties or unlinkability levels. For a given score $s$, $LR(s)$ is defined as

$$LR(s) = \frac{p(s|H_m)}{p(s|H_{nm})} \tag{7}$$

In particular, two different cases can be defined based on $LR(s)$:

- If $LR(s) \leq 1$, we can state that it is more likely that both templates belong to non-mated instances, thereby making the templates unlinkable for those score values. Therefore, we will have $\mathrm{D}_{\leftrightarrow}(s) = 0$.
  Bear in mind that a system is considered to be linkable if it allows determining, with some certainty, that two templates come from the same person. In the case of $LR(s) \leq 1$, a potential attacker knows, with some certainty, that both templates do *not* belong to the same subject and therefore he cannot link them. That is why for those score values the system is considered to be unlinkable, i.e., $\mathrm{D}_{\leftrightarrow}(s) = 0$.
- If $LR(s) > 1$, we can state that it is more likely that both templates belong to the same instance, thereby making the templates somewhat linkable for those score values. In fact, the higher $LR(s)$, the more linkable the templates are. As a consequence, we will define an increasing value $\mathrm{D}_{\leftrightarrow}(s) \in (0, 1]$, with higher values for more linkable templates (i.e., the higher $LR(s)$, the closer $\mathrm{D}_{\leftrightarrow}(s)$ to 1).

Keeping those remarks in mind, we define $\mathrm{D}_{\leftrightarrow}(s)$ as a function of $s$ and its corresponding $LR(s)$. Since $LR(s)$ yields values in the range $[0, \infty)$, in order to obtain the desired measure in the range [0, 1], we perform a two step normalisation. In the first step, we normalise $LR(s) - 1$ to the range [0.5, 1] with a sigmoid function. Then, we subtract 0.5 and multiply by 2 to map that interval to [0, 1]. Therefore, we can finally define $\mathrm{D}_{\leftrightarrow}(s)$ as

$$\mathrm{D}_{\leftrightarrow}(s) = \begin{cases} 0 & \text{if } LR(s) \leq 1 \\ 2 \cdot \left( \left(1 + e^{-(LR(s)-1)}\right)^{-1} - 0.5 \right) & \text{if } LR(s) > 1 \end{cases} \tag{8}$$

By the definition of the sigmoid function,

$$\left(1 + e^{-(LR(s)-1)}\right)^{-1} \to 0.5 \qquad \text{when } LR(s) \to 1 \tag{9}$$

$$\left(1 + e^{-(LR(s)-1)}\right)^{-1} \to 1 \qquad \text{when } LR(s) \to \infty \tag{10}$$

Therefore, the second step of the normalisation moves the range of values of $\mathrm{D}_{\leftrightarrow}(s)$ from [0.5, 1] to [0, 1], as desired. Additionally, the proposed metric is continuous for $LR(s) = 1$.

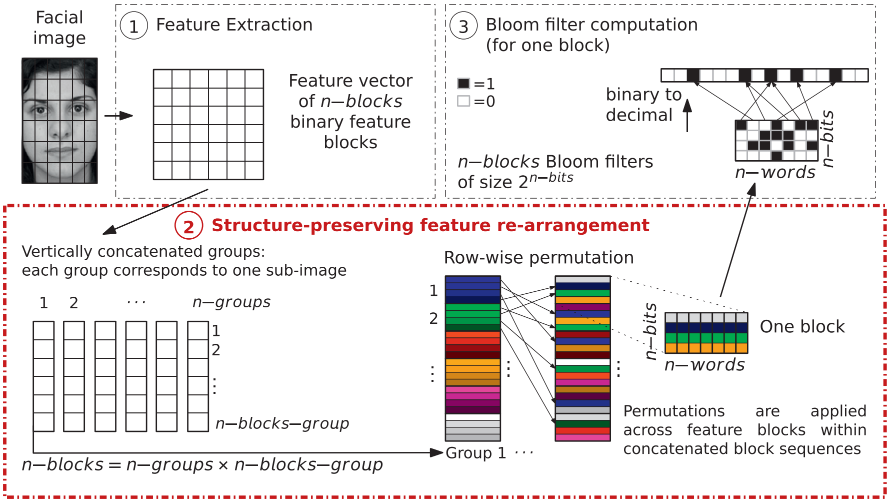

Facial image
1 Feature Extraction
Feature vector of n−blocks binary feature blocks
3 Bloom filter computation (for one block)
■=1
□=0
binary to decimal
n−blocks Bloom filters of size 2^{n−bits}
n−words
n−bits
2 Structure-preserving feature re-arrangement
Vertically concatenated groups: each group corresponds to one sub-image
1 2 ··· n−groups
1
2
⋮
n−blocks−group
n−blocks = n−groups × n−blocks−group
Row-wise permutation
1
2
⋮
Group 1 ···
One block
n−bits
n−words
Permutations are applied across feature blocks within concatenated block sequences

**Fig. 2.** System overview: (1) a binary feature vector consisting of $n{-}blocks$ binary feature blocks is extracted; (2) the entire set of blocks is disposed into $n{-}groups$ vertically concatenated groups consisting of $n{-}blocks{-}group$ blocks, and structure-preserving feature re-arrangement is applied; (3) a total number of $n{-}blocks$ Bloom filters is extracted (one for each transformed feature block).

As described previously, it is also useful to have an estimation of the *unlinkability of the whole system* (and not for every single score). For this purpose, we define $\mathrm{D}_{\leftrightarrow}^{sys}$ as the partial area under the curve $\mathrm{D}_{\leftrightarrow}(s)$, normalised by $p(s|H_m)$ in order to get values in [0, 1], and computed on the whole score range (i.e., $[s_{min}, s_{max}]$):

$$\mathrm{D}_{\leftrightarrow}^{sys} = \int_{s_{min}}^{s_{max}} \mathrm{D}_{\leftrightarrow}(s) \cdot p(s|H_m) \mathrm{d}s \tag{11}$$

Since $\mathrm{D}_{\leftrightarrow}(s) \in [0, 1]$ and $\int_{s_{min}}^{s_{max}} p(s|H_m) = 1$, the global linkability measure yields values in [0, 1], as desired. More specifically, the final value of $\mathrm{D}_{\leftrightarrow}^{sys}$ depends on: *i*) the range of scores where the system is linkable; *ii*) how linkable the system is in that range of scores; and *iii*) how probable it is that such scores are produced.

Let us now evaluate the scenarios shown in Fig. 1 with the proposed measures For the fully unlinkable scenario shown in Fig. 1a, $\mathrm{D}_{\leftrightarrow}(s) = 0$ for all scores, thus yielding the desired value $\mathrm{D}_{\leftrightarrow}^{sys} = 0$.

For the semi-unlinkable scenario shown in Fig. 1b, we may observe that $\mathrm{D}_{\leftrightarrow}(s) = 0$ for all scores where $H_{nm}$ holds (i.e., $s \notin [0.79, 0.81]$). Additionally, $\mathrm{D}_{\leftrightarrow}(s)$ reaches a maximum of 0.2 for a score value of 0, where the *LR* is the highest ($LR(0) \sim 1.5$) and we can thus assume with the highest certainty that both templates conceal the same identity. Overall, we obtain $\mathrm{D}_{\leftrightarrow}^{sys} = 0.14$, reflecting the fact that only for a small range of scores we could link templates.

For the semi-linkable scenario shown in Fig. 1c, we observe that $\mathrm{D}_{\leftrightarrow}(s) = 1$ for scores out of [0.79, 0.81], where $p(s|H_{nm}) = 0$, and we can assume with almost all certainty that the compared templates conceal the same instance. As a consequence, we obtain a higher value for $\mathrm{D}_{\leftrightarrow}^{sys} = 0.32$, with respect to the semi-unlinkable scenario.

For the fully linkable scenario shown in Fig. 1d, we observe that $\mathrm{D}_{\leftrightarrow}(s) = 1$ where only the *Mated instances* distribution is non-null (i.e., $s \in [0.55, 0.65]$), since templates are fully linkable in such range. On the other hand, $\mathrm{D}_{\leftrightarrow}(s) = 0$ in any other place. Therefore the system as a whole is fully linkable, as it holds that $\mathrm{D}_{\leftrightarrow}^{sys} = 1$, as desired.

Finally, it should be noted that such unlinkability analysis is not sufficient to ensure the cross-matching resistance of protected templates, since the robustness against specifically designed attacks has to be analysed as well [36]. To that end, the aforementioned distributions will be estimated not only for the dissimilarity scores computed by the biometrics system, but also for other distance measures appropriate for the cross-matching attack at hand (e.g., Hamming Distance, Hamming Weight difference), and analysed in the same manner.

## 4. Improved system

An overview of the processing chain of the proposed improved Bloom filter-based template protection scheme is depicted in Fig. 2. In contrast to the original concept, an additional processing step, referred to as *Structure-preserving feature re-arrangement*, is introduced. Hence, the improved scheme comprises three key components:

1. *Feature extraction*: in the first step, an unprotected two-dimensional binary feature vector is extracted from an image, e.g. pre-aligned facial image. In the same way as in the original concept, the binary feature vector is divided into $n{-}blocks$ blocks of size $n{-}bits \times n{-}words$ bits, as shown as part of Fig. 2.
2. *Structure-preserving feature re-arrangement*: the goal of this processing step is to dissipate the statistical composition of the biometric feature vector. In order to maintain recognition performance, a certain structure of words in feature blocks

has to be retained. Otherwise, stability of discriminative words is lost prior to the computation of Bloom filters. In order to reach a balance between biometric performance and diffusion of feature vectors, we first re-group $n{-}blocks$ blocks into a set of $n{-}groups$ concatenated groups consisting of $n{-}blocks{-}group$ blocks, $n{-}blocks = n{-}groups \times n{-}blocks{-}group$, see Fig. 2. Hence, one group of feature blocks corresponds to a sub-image of the captured sample.

Within such groups of blocks, a *row-wise permutation* ($perm$) is performed: for each of the $n{-}groups$ sets, the rows of the vertical concatenation of corresponding $n{-}blocks{-}group$ blocks are permuted. Note that a permutation of columns would not cause any change in the resulting Bloom filters. Since horizontal neighbourhoods of bits within rows persist, this sub-step prevents from a potential loss of discriminative power of resulting feature blocks. The dissipation of rows among groups of blocks significantly improves the information diffusion and prevents block-based attacks. In case of a permutation within feature blocks, a potential attacker, which has full knowledge of the employed permutation key (full disclosure model), would be able to revert Bloom filters to feature blocks separately after applying the reverse permutation, which involves an arrangement of $|\mathbf{b}|$ words to a block of length $n{-}words$ with $n{-}seq$ possible sequences of words. However, applying a inverse permutation across a group of blocks prior to reverting Bloom filters to feature blocks is not feasible, since without loss of generality, the number of activated bits in Bloom filters of feature blocks of one group differs. This means that, after applying the correct inverse permutation adjacencies of bits forming each word are potentially lost. As a consequence, one out of $n{-}seq$ sequences would have to be guessed for each of the $n{-}blocks{-}group$ blocks of a group, prior to applying the inverse permutation. Moreover, the re-grouping of feature blocks increases the size of the key space for the applied permutation.

3. *Bloom filter computation*: in the final step one Bloom filter is computed from each of the $n{-}blocks$ blocks, such that the final protected template $\mathbf{C}$ consist of $n{-}blocks$ Bloom filters of size $2^{n-bits}$. An example of this processing step for a single feature block is shown as part of Fig. 2.

Despite the proposed structure-preserving feature re-arrangement, a random shuffling of bits would fulfil the task of dissipating the statistical composition of the biometric feature vector. However, such an approach significantly affects biometric performance, as will be shown in the experiments. Alternatively, XOR-ing the entire feature vector with a randomly generated binary vector of the same size (one-time pad) could be considered. However, while such an approach would achieve sufficiently large key spaces, block-based attacks could be employed in a scenario where an attacker has full knowledge of the applied key, since biometric information would not be dispersed across feature blocks prior to the bloom filter computation.

The level of unlinkability and irreversibility achieved by the proposed system will be influenced by the size of the key space, $|\mathbf{T}|$, of the considered structure-preserving feature re-arrangement. Two facts should be taken into account for the computation of $|\mathbf{T}|$: *i*) the dimensions of feature blocks and concatenated groups of blocks to which the *perm* transform is applied, and *ii*) the number of feature blocks and concatenated groups of blocks, since different keys are applied for each of these. In our particular approach, we are carrying out $n{-}groups$ different permutations (one for each group of blocks) of $n{-}bits \times B$ rows. Therefore, for each permutation we have $(n{-}bits \times B)!$ different keys resulting in,

$$|\mathbf{T}| = (n{-}bits \times B)!^{\,n-groups} \tag{12}$$

In contrast to the original approach [33], key space sizes of the proposed structure-preserving feature re-arrangement are large enough to prevent brute force cross-matching attacks, as will be shown in the experiments.

## 5. Potential attacks

An eventual attacker may take advantage of certain statistical properties or weaknesses of the template protection scheme. For this reason, the robustness of the proposed improved system needs to be analysed with respect to already proposed as well as foreseeable attacks. To that end, two different adversary models will be considered:

- *Advanced model*: In this model, the adversary has the full knowledge of the algorithms used for template extraction, template protection and comparison, following Kerckhoffs principles [20]. In addition, the adversary is capable of executing part of or all sub-modules of the system that make use of the secret keys, while the adversary knows none of the secrets.
- *Full disclosure model*: this model is the advanced model augmented by disclosing the secret keys to the adversary.

It should be noted that it is implied that a successful attack on the irreversibility property of a template protection system also breaks unlinkability, i.e. enables cross-matching.

### *5.1. Brute force attack*

A brute-force cross-matching attack on the original concept of Bloom filter-based template protection has been proposed in [7]. Let $\mathbf{M}$ be a biometric datum which is protected applying two different secret keys $\mathbf{T}_1$ and $\mathbf{T}_2$ resulting in $\mathbf{C}^1 = PIE(\mathbf{M}, \mathbf{T}_1)$ and $\mathbf{C}^2 = PIE(\mathbf{M}, \mathbf{T}_2)$. Since the indexes of the resulting sets of Bloom filters $\mathbf{C}^1 = \{\mathbf{b}_1^1, \mathbf{b}_2^1, \ldots, \mathbf{b}_{n-blocks}^1\}$ and $\mathbf{C}^2 = \{\mathbf{b}_1^2, \mathbf{b}_2^2, \ldots, \mathbf{b}_{n-blocks}^2\}$ are visible to an attacker, the following strategy can be employed to cross-match $\mathbf{C}^1$ and $\mathbf{C}^2$. Each index of one of the two associated Bloom filters is XORed with every possible secret $\mathbf{T}^* \in \{0,1\}^{n-bits}$ and it is checked whether $\mathbf{b}_i^1[j] = \mathbf{b}_i^2[j] \oplus \mathbf{T}^*$, $j = 0, \ldots, 2^{n-bits} - 1$, holds for all non-zero indexes. This attack can also be applied if $\mathbf{C}^1$ and $\mathbf{C}^2$

are generated from different biometric inputs of the same subject, by searching for a $\mathbf{T}^*$ which yields a minimum dissimilarity score (*DS*) between $\mathbf{C}^1$ and $\mathbf{C}^2$. In case binary blocks are large enough, the brute-force search will also succeed if different keys are used for different blocks.

### 5.2. Reconstruction attack

Given a protected template $\mathbf{C}$, the goal of this attack is to reconstruct a biometric datum $\mathbf{M}'$, which is close to the original biometric input $\mathbf{M}$, i.e. the attack can be employed to break irreversibility and unlinkability. Given one Bloom filter $\mathbf{b}$, for each activated index $i = 1, \dots, |\mathbf{b}|$, the corresponding word $\mathbf{s}_i$ is reconstructed. The entire feature block is reconstructed as one single word repeated $n\!-\!words$ times, where that word represents the bit-wise average of the $|\mathbf{b}|$ reconstructed words, i.e. in the final feature word $\mathbf{s}$, a given bit is activated iff it was activated at least $|\mathbf{b}|/2$ times. In [7] this attack was applied to the original iris-based scheme proposed in [33] without applying any secret keys. It is shown that, even though the reconstructed iris-codes have not a realistic appearance, the *HD* between them and the original iris-codes is below the threshold set at FMR $= 10^{-4}$, thus positively matching the original iris-codes and granting access to eventual impostors.

### 5.3. Hamming weights attack

An efficient cross-matching attack on the original proposal of Bloom filter-based template protection is presented in [15]. This attack takes advantage of the fact that if *W* different words appear within one processed binary block, *W* different bits will be set to one in the corresponding Bloom filter: the proposed XOR represents a linear mapping, i.e. no collisions will occur. Let us assume that one biometric input $\mathbf{M}$ is protected applying two different secret keys $\mathbf{T}_1$ and $\mathbf{T}_2$, resulting in $\mathbf{C}^1 = PIE(\mathbf{M}, \mathbf{T}_1)$ and $\mathbf{C}^2 = PIE(\mathbf{M}, \mathbf{T}_2)$. This means that, regardless of the values of $\mathbf{T}_1$ and $\mathbf{T}_2$, the Hamming Weights (*HW*) of $\mathbf{C}_1$ and $\mathbf{C}_2$ will be identical, $|\mathbf{C}^1| = |\mathbf{C}^2|$, since $|\mathbf{b}_1^1| = |\mathbf{b}_1^2|, |\mathbf{b}_2^1| = |\mathbf{b}_2^2|, \dots, |\mathbf{b}_{n-blocks}^1| = |\mathbf{b}_{n-blocks}^2|$. Based on a theoretical analysis for the setting proposed in [33], the authors report that, in the worst case scenario, this trivial cross-matching attack succeeds with a probability of at least 96%.

### 5.4. Exploiting the XOR-operation

In the original concept of Bloom filter-based template protection, the application of a XOR operation represents a linear transform, which is applied to each word of each binary block. Let us assume that one biometric sample $\mathbf{M}$ is protected applying two different secret keys $\mathbf{T}_1$ and $\mathbf{T}_2$, resulting in $\mathbf{C}^1$ and $\mathbf{C}^2$, respectively. An attacker can now analyse bit-vectors consisting of the *i*th indexes of all Bloom filters in $\mathbf{C}^1$ and search for an identical vector in $\mathbf{C}^2$. Since the same secret key is applied to generate all Bloom filters of one protected template, for each vector $(\mathbf{b}_1^1[i], \mathbf{b}_2^1[i], \dots, \mathbf{b}_{n-blocks}^1[i])$, $i = 0, \dots, 2^{n-bits} - 1$, there will be an identical vector $(\mathbf{b}_1^2[j], \mathbf{b}_2^2[j], \dots, \mathbf{b}_{n-blocks}^2[j])$, $j = 0, \dots, 2^{n-bits} - 1$. It is important to note that the mapping between all vectors of $\mathbf{C}_1$ and all vectors of $\mathbf{C}_2$ is bijective. In other words, the XOR operation produces a linear shift of indexes within Bloom filters which is identical for each block. This fact can be exploited by an attacker to cross-match two protected templates at reduced computational cost, compared to the brute force attack.

Moreover, this attack can be extended to link protected templates generated from different biometric samples $\mathbf{M}_1 \neq \mathbf{M}_2$ of the same instance. In this case, given $\mathbf{C}^1$ and $\mathbf{C}^2$, the attacker would search for corresponding bit vectors exhibiting a minimum *HD*, thus obtaining a permuted template $\mathbf{C}^{2'}$. The final decision on whether $\mathbf{C}^1$ and $\mathbf{C}^2$ belong to the same subject will be based on the *HD* between the first and the permuted templates, i.e., $HD(\mathbf{C}^1, \mathbf{C}^{2'})$.

## 6. Experimental setup

### 6.1. Database

In order to make the present study reproducible and comparable to future research, experiments are carried out on the widely used public face subcorpus of the Desktop Dataset (DS2) of the BioSecure Multimodal Database[2] [27]. The face subset used in this work includes four frontal images of 210 subjects, captured in two time-spaced acquisition sessions (two images per session), with an homogeneous grey background and using a reflex digital camera (8.2 MP resolution) without flash ($210 \times 4 = 840$ face samples). Eyes were automatically annotated applying VeriLook SDK 4.0, provided by Neurotechnology.[3]

### 6.2. Face verification system

The face verification system that serves as baseline for the proposed approach is an implementation of the Local Gabor Binary Pattern Histogram Sequences (LGBPHS) algorithm [40], a state-of-the-art system robust to illumination changes. In a fair benchmark among four state-of-the-art algorithms for face recognition established in [14], using the same databases and

[2] Publicly available at http://biosecure.it-sudparis.eu/AB

[3] http://www.neurotechnology.com/verilook.html

protocols, LGBPHS achieved a top performance. Feature extraction is applied in a block-wise manner, i.e. the facial image is divided into $n-groups$ non-overlapping sub-images, from which spectral histograms are computed and concatenated to form the final template. For more details on the employed feature extraction, the reader is referred to [40].

Experiments are run using Bob[4] [4], a free signal and image processing toolbox, which includes a library with implementations of several face verification algorithms – the Facereclib [14]. We used its implementation of LGBPHS, considering only the central $n-groups = 32$ sub-images for verification purposes. A single configuration for the Bloom filter extraction is selected for the experiments so that the study is kept within a reasonable length: $n-bits = 5$ and $n-words = 15$, which shows a good balance between irreversibility and biometric performance. In [12], further possible configurations can be found. Therefore, in this particular implementation, $32 \times 40 = 1280$ histograms of 60 bins are computed and concatenated. Each sub-image is then further divided into $n-blocks-group = (40/n-bits) \times (60/n-words) = 32$ blocks, in order to compute a total number of $n-blocks = n-groups \times n-blocks-group = 32 \times (40/n-bits) \times (60/n-words) = 1024$ Bloom filters for the final protected template. For the particular structure-preserving feature re-arrangement proposed, each sub-image will be regarded as one of the block sets to which the *perm* is applied.

### 6.3. Experimental protocol

The evaluation protocol is designed to estimate: *i*) to what extent the proposed approach meets the requirements of template protection systems defined in [16], and also *ii*) what is the improvement achieved with respect to the original system proposed in [12], especially in terms of unlinkability, which was one of the main limitations of the previous approach. Therefore, the protocol comprises four different evaluations:

*Performance evaluation*: the first question to analyse is the impact of the proposed improvements on the biometric performance of the system. Therefore, the performance variation between the baseline system and the protected system is evaluated in the first set of experiments. In accordance with ISO/IEC IS 19795-1 [17], performance is evaluated in terms of False Non-Match Rate (FNMR), i.e. the proportion of genuine verification attempts rejected, and False Match Rate (FMR), i.e. the proportion of zero-effort impostor attempts accepted as genuine samples. In this context the Equal Error Rate (EER) is defined as the point where FNMR = FMR. Biometric performance is evaluated under the stolen-token scenario, i.e. one single randomly generated key is employed for each configuration of the improved system.

*Irreversibility analysis*: once the performance has been evaluated, the irreversibility provided by the proposed improved Bloom filter-based template protection system is analysed. To that end, two different aspects will be considered: (*i*) the success probability of guessing the correct original template, and (*ii*) given a protected template, the probability that the corresponding unprotected template will be reconstructed applying the reconstruction attack. In this last case, the quality of the reconstructed template is estimated by comparing it to the corresponding original binary feature vector in terms of *HD*. In the advanced model, attacks on irreversibility also involve guessing the inverse transforms applied during the structure-preserving feature re-arrangement. In order to analyse the irreversibility achieved by the proposed method, the resulting score distributions will be compared to that of random impostors.

*Unlinkability analysis*: in order to assess whether the improved system proposed in the present work meets the unlinkability requirement, the methodology defined in Section 3 will be used to analyse and compare the original [12] and improved schemes.

*Robustness to potential attacks*: finally, all proposed cross-matching attacks are applied to the both systems.

## 7. Experimental evaluation

### 7.1. Performance evaluation

The unprotected baseline system achieves an EER of 6.25%. Biometric performance obtained for different configurations of the improved system is summarized in Fig. 3. Regarding the improved system, denoted as *perm*, almost no change in verification performance is observed (relative change as low as 1.3%), which confirms that the proposed transforms retain the structure of the original unprotected feature vectors. Moreover, as shown in Fig. 3, characteristics of detection error trade-off (DET) curves are similar to that of the baseline system, hence confirming that the improved system's accuracy is preserved. In contrast, a random shuffling of bits within according groups of blocks causes an increase of the EER to over 40%.

### 7.2. Irreversibility analysis

For the improved face-based template protection scheme, the average number of bits set to one for a given Bloom filter, denoted as $|\overline{\mathbf{b}}|$, and the corresponding average number of re-mapped words $\overline{rm-rate}$, $\overline{rm-rate} = 1 - |\overline{\mathbf{b}}|/n-words$, are empirically obtained from the protected templates of all samples in the database. Based on these values, the average number of possible sequences $\overline{n-seq}$ resulting in a single Bloom filter, defined in Eq. 1, is raised to $n-blocks$, the number of Bloom

[4] Publicly available at http://idiap.github.io/bob/

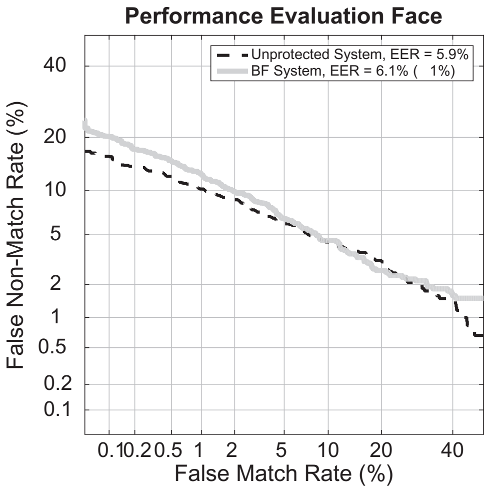


**Fig. 3.** Performance analysis: comparison of DET curves for the improved protected system (solid line) and the unprotected baseline system (dashed line).

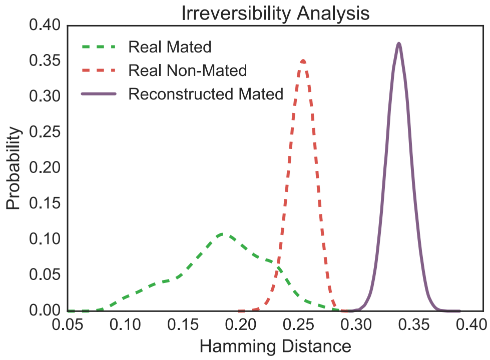


**Fig. 4.** Irreversibility analysis: *HD*-based score distributions between the reconstructed and the original unprotected templates, compared to the genuine and random impostor scores between real unprotected templates.

filters forming protected templates, in order to estimate the entire inverse image set of the protected template prior to the Bloom filter computation.

Therefore, given a protected template, the success probability of guessing the corresponding unprotected feature vector is estimated as $\overline{n-seq}^{-n-blocks}$ for the full disclosure model, where **T** is known to the adversary. In the case of the advanced model, an attacker would further have to guess the employed key **T**, i.e. the success probability of guessing unprotected feature vectors is calculated as $\overline{n-seq}^{-n-blocks} \times |\mathbf{T}|^{-1}$, which for some configurations is significantly smaller than directly guessing the feature vector of size $n-blocks \times n-words \times n-bits = 76,800$. Table 1 summarizes the results obtained for the improved system with respect to the level of irreversibility provided, where key space size is estimated as follows, see Eq. 12,

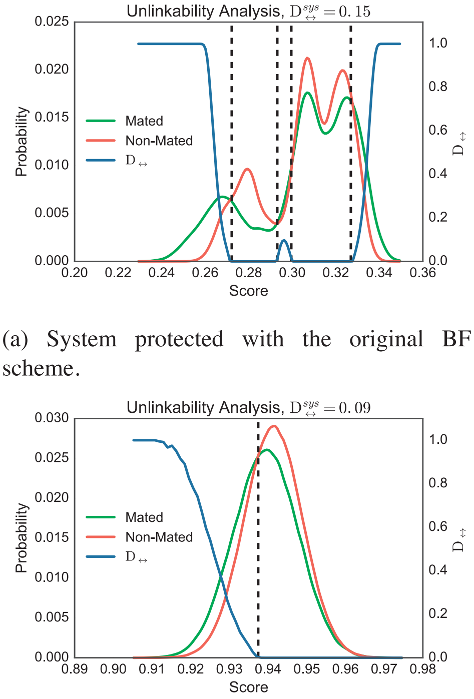
$$|\mathbf{T}| = (5 \times 32)!^{32} \approx 2^{30,261} \tag{13}$$

It is important to note that estimations yield lower bounds for success probabilities, since these refer to the probability of guessing the correct original template and not a template which is close to the original one. Still, even in case the attacker is in possession of protected templates and their corresponding keys it is still not possible to directly revert the protected template to the original feature vector. As it may be observed, the success probability of guessing the correct unprotected template is below $2^{-40,000}$ ($\sim 10^{-12,000}$).

Focusing on the reconstruction attack proposed in [7], *HD*-based distributions between the original unprotected templates and the ones obtained with the suggested reconstruction attack are depicted in Fig. 4, where only the full disclosure model has been considered. While in the original system proposed in [12] the reconstructed templates distribution (solid line) overlapped with the genuine scores distribution (dashed green line), now the impostor *HDs* (dashed red line) are even lower than those obtained with the reconstructed templates for the improved system. As a consequence, the reconstructed templates are no longer accepted into the system. We can hence conclude that, even for the full disclosure model, the improved system does not allow an efficient reconstruction of templates close to the original ones. Furthermore, yielding

**Table 1**
Irreversibility evaluation: average number of bits set to one per Bloom filter, average percentage of re-mapped words, average number of possible sequences per block, and success probabilities for guessing original unprotected templates.

| $\lvert\bar{\mathbf{b}}\rvert$ | $\overline{rm-rate}$ (%) | $\overline{n-seq}$ | Success probability | |
|---|---|---|---|---|
| | | | Advanced | Full disclosure |
| 6.56 | 56.3 | $2^{40}$ | $2^{-71,221}$ | $2^{-40,960}$ |



(a) System protected with the original BF scheme.

(b) System protected with the improved BF scheme.

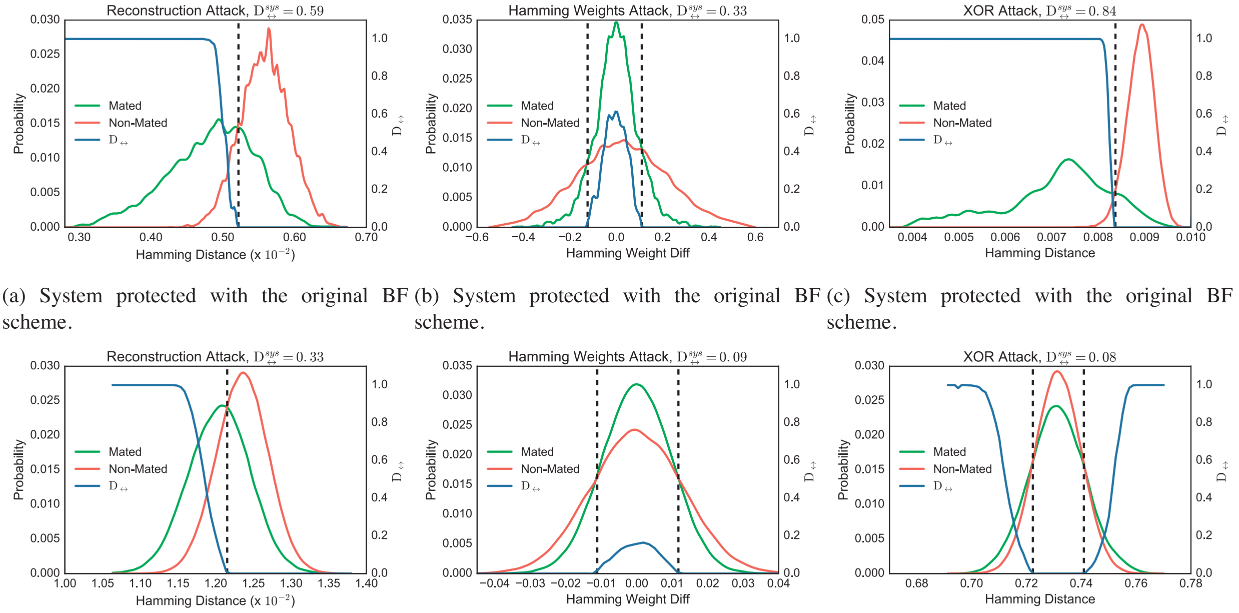
**Fig. 5.** Unlinkability analysis: scores distributions for comparisons of protected templates generated with $n-keys = 10$ different keys for the original scheme (top) and the improved system (bottom) The dashed black line represents $LR(s) = 1$.

*HDs* higher than the random impostor distances implies that also the security of the system is enhanced, since access will not be granted to such reconstructed templates.

### 7.3. Unlinkability analysis

In order to assess the level of unlinkability provided by the improved system, we will follow the protocol established in Section 3. To that end, the two score distributions (i.e., *Mated instances* and *Non-mated instances*) are compared in Fig. 5 for sets of $n-keys = 10$ secret keys. $D_{\leftrightarrow}(s)$ and $D_{\leftrightarrow}^{sys}$ are also depicted in the same figure. In this and the subsequent figures, " Original system" refers to the original Bloom filter based template protection scheme presented in [12,33], and "Improved system" to the scheme proposed in this article.

As it may be observed in Fig. 5, the distributions obtained for the improved system (Fig. 5b) overlap to a bigger extent than those corresponding to the original system (Fig. 5a). In particular, the linkable, and thereby vulnerable, range of scores for the original system (i.e., those $s$ for which $LR(s) > 1$, and, according to Eq. 8, $D_{\leftrightarrow}(s) > 0$) is bigger and has a higher probability mass than that of the improved system. As a consequence, $D_{\leftrightarrow}^{sys}$ is 67% lower than that of the original system, yielding a value as low as 0.09. We may thus conclude that templates are almost unlinkable when compared in terms of their dissimilarity scores.



(a) System protected with the original BF scheme.

(b) System protected with the original BF scheme.

(c) System protected with the original BF scheme.

(d) System protected with the improved BF scheme.

(e) System protected with the improved BF scheme.

(f) System protected with the improved BF scheme.

**Fig. 6.** Robustness to cross-matching attacks: distributions for the analysis of three different cross-matching attacks for the original scheme (top) and the improved system (bottom). The dashed black line represents $LR(s) = 1$.

### 7.4. Robustness to proposed cross-matching attacks

In order to analyse the uniformity of the templates, the entropy of the protected templates is compared to that of the unprotected templates. The entropy, $E$, is defined as, $E = -\sum p \log p$, where $p$ is the probability of occurrence of a given value. In our particular case, the distribution of bits set to one is first estimated over the whole database, yielding the $p$ probabilities. Then the entropy of those distributions is computed, yielding $E = 4.01$ for feature vectors of the original unprotected system, and $E = 4.08$ for the protected templates. Since the entropy is maintained between both systems, no additional correlations are introduced by the protection scheme and therefore they cannot be exploited by eventual statistical attacks.

#### 7.4.1. Brute force attack

In the advanced model the efficiency of a brute force cross-matching attack depends on the size of the key space: on average, an attacker needs to guess correct sequences of words within feature blocks as well as the key in order to succeed. The average success probability of this attack can be thus estimated as $2(\overline{n-seq}^{-n-blocks} \times |\mathbf{T}|)^{-1}$. Since the suggested structure-preserving feature transforms obscure rows among $nGroups = 32$ groups of binary blocks, the success rate of cross-matching attacks may be increased to $2(\overline{n-seq}^{-n-blocks/n-groups} \times |\mathbf{T}|)^{-1}$, in case the attack is applied simultaneously to each group of blocks. However, even if a brute force cross-matching attack is parallelized for groups of blocks, success rates for the improved system remain rather low. We thus conclude that brute force cross-matching attacks are computationally infeasible.

In the full disclosure model cross-matching would involve guessing the inversion of the Bloom filter-based transform prior to performing the inverse structure-preserving feature re-arrangement, hence, success rates increase to $2(\overline{n-seq}^{-n-blocks})$. For parallelized group-based attacks success rates further increase to $2(\overline{n-seq}^{-n-blocks/n-groups})$, yielding success rates below $2^{-1279}$. It should be noted that, if block-based transforms such as the XOR with one-time pad were applied, success rates would increase to $2^{-39}$ in case cross-matching is performed simultaneously for each block.

#### 7.4.2. Reconstruction attack

In case of cross-matching, the aim of this attack is to revert two protected templates and link them. Given the low success probabilities estimated for brute force attacks, in this case we restrict the analysis to the full disclosure model. As in the initial unlinkability analysis, the distributions of the *HD*s between the reconstructed unprotected templates generated from the *Mated instances* or *Non-mated instances*, are depicted in Figs. 6a and 6d, for the original and improved systems, respectively.

Similar to the unlinkability analysis, the distributions *Mated instances* and *Non-mated instances* overlap to a bigger extent for the improved system, reducing the final $D_{\leftrightarrow}^{sys}$ in 70%, from 0.59 to 0.33. As a consequence, even if the system is more vulnerable to this attack than to the analysis of plain scores under a normal operation mode, for which $\mathrm{D}_{\leftrightarrow}^{sys} = 0.09$ (Fig. 5b), we may conclude that the templates' robustness to this cross-matching attack has been considerably improved.

In addition, it should be noted that this attack assumes the highest amount of knowledge on the attacker, who is in possession of the secret keys used by the system. Therefore, the $\mathrm{D}_{\leftrightarrow}^{sys}$ reaches its highest value for all the cross-matching attacks analysed.

### 7.4.3. *Hamming weights attack*

The *HW*s of the protected templates might be used to cross-match templates generated with different keys. The distributions of the differences in *HW*s between protected templates generated from the *Mated instances* or *Non-mated instances*, are depicted in Figs. 6b and 6e, for the original and improved systems, respectively. As we may observe, in both cases all distributions are centred on the same value, zero. However, for the original system $LR(s)$ is higher in the linkable range of scores (i.e., $s$ such that $LR(s) > 1$). As a consequence, in that range $\mathrm{D}_{\leftrightarrow}$ is also higher, reflecting the higher vulnerability of the original system to this attack.

Additionally, while for the original system $\mathrm{D}_{\leftrightarrow}^{sys} = 0.33$, twice as large as under a normal operation mode (Fig. 5a), in the improved system $\mathrm{D}_{\leftrightarrow}^{sys} = 0.09$ (i.e., same value obtained for the improved system under normal operational conditions with no attack shown in Fig. 5b), reducing the linkability in over 250%. We can hence conclude that templates are robust to this cross-matching attack.

### 7.4.4. *Exploiting the XOR-operation*

The XOR operation proposed in the original concept of Bloom filter-based template protection might be exploited to carry out a cross-match attack. To apply this attack, we need to compute *HD*s between optimally re-permuted protected templates. Then, the distributions of such distances, generated from *Mated instances* or *Non-mated instances*, are depicted in Figs. 6c and 6f, for the original and improved systems, respectively. As can be observed, the original system is highly vulnerable to this attack: both distributions are easily separable, except for a small range of scores, hence yielding $\mathrm{D}_{\leftrightarrow}^{sys} = 0.84$. On the other hand, for the improved system, only the tails of the *Mated instances* distributions are slightly heavier, thus showing values close to 1 for $\mathrm{D}_{\leftrightarrow}(s)$. This means that the templates yielding those distances are more likely to belong to the same instance. However, since the scores presenting high $D_{\leftrightarrow}(s)$ values (i.e., the distribution tails) are very unlikely to happen, the final unlinkability value achieved for the system is very low, $\mathrm{D}_{\leftrightarrow}^{sys} = 0.08$, which is over ten times smaller than that of the original system, and below the one obtained for the improved system working on normal operation conditions with no attack (see Section 7.3). Therefore, we may conclude that, unlike the original system, the improved system is robust to cross matching attacks based on the XOR-Operation.

## 8. Conclusions

Given the wide deployment of biometric recognition systems for everyday tasks, such as withdrawals in ATMs or border crossing, the protection of the privacy of the subject has become a key issue of this technology. The development of new template protection techniques and the thorough evaluation of their irreversibility and unlinkability properties is therefore of the utmost importance.

In the present work, we first introduce a new framework, based on likelihood ratios, for the unlinkability evaluation of protected templates. Then, we present an improved Bloom filter-based template protection scheme by proposing an easily integratable processing step, referred to as structure-preserving feature re-arrangement, for the purpose of dissipating the statistical composition of the biometric feature vector. At the same time, the structure of discriminative feature parts is retained, i.e. biometric performance rates of the corresponding unprotected recognition system is maintained in the stolen-token scenario.

In a fully reproducible experimental study, which is conducted for a facial Bloom filter-based protection scheme, irreversibility and unlinkability are confirmed, considering an advanced adversary model, as well as a full disclosure adversary model, where a potential attacker is in possession of secret keys. In particular, focusing on the full disclosure adversary model the chance of reverting a protected template and cross-matching two protected templates is below $10^{-40,960}$, which is far below the FMR of conventional biometric (template protection) systems. It is important to note, that the full disclosure adversary model is commonly neglected with respect to attacks on irreversibility or unlinkability. It is questionable whether the vast majority of current cancelable biometric schemes would resist in such a scenario, while in a biometric cryptosystem, knowledge of the secret key implies full exposure of the protected biometric data.

In addition, the template size is reduced with respect to the original unprotected template, and verification is carried out in a fast efficient manner, thus allowing the deployment of the proposed system in real-time applications. For applications which may need a compact key-space we suggest to employ shorter keys, e.g. 128 bit, as input of random number generators in order to generate keys of a required size.

The proposed framework for unlinkability analysis only takes into account one-to-one comparisons, when the attacker is in possession of two protected templates and wants to decide whether they belong to the same subject. As future work lines, we will further investigate the more general case when the attacker can compare a single template with a database of $N$ different templates and decide whether any of them conceal the same identity.

## Acknowledgements

This work has been partially supported by projects CogniMetrics (TEC2015-70627-R) from Spanish MINECO/FEDER, FIDELITY (FP7-SEC-284862) and BEAT (FP7-SEC-284989) from EU, the Center for Advanced Security Research Darmstadt (CASED), and *CECABANK*. M. Gomez–Barrero is supported by a FPU Fellowship from Spanish MECD. This work was conducted during a research stay at NISlab.

## References

[1] N. Abe, S. Yamada, T. Shinzaki, Irreversible fingerprint template using minutiae relation code with bloom filter, in: Proceedings of International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2015.
[2] S. Abuguba, M.M. Milosavljevic, N. Macek, An efficient approach to generating cryptographic keys from face and iris biometrics fused at the feature level, Int. J. Comput. Sci. Netw. Security (IJCSNS) 15 (6) (2015) 6.
[3] G. Amirthalingam, G. Radhamani, Multimodal biometric cryptosystem for face and ear recognition based on fuzzy vault, Res. J. Appl. Sci. Eng. Technol. 7 (20) (2014) 4211–4219.
[4] A. Anjos, L.E. Shafey, et al., Bob: a free signal processing and machine learning toolbox for researchers, in: Proceedings of ACM MM, 2012, pp. 1449–1452.
[5] B. Bloom, Space/time tradeoffs in hash coding with allowable errors, Commun. ACM 13 (7) (1970) 422–426.
[6] T. Boult, Robust distance measures for face-recognition supporting revocable biometric tokens, in: Proceedings of ICAFGR, 2006, pp. 560–566.
[7] J. Bringer, C. Morel, C. Rathgeb, Security analysis of bloom filter-based iris biometric template protection, in: Proceedings of ICB, 2015, pp. 527–534.
[8] P. Campisi (Ed.), Security and Privacy in Biometrics, Springer, 2013.
[9] A.M. Canuto, F. Pintro, M.C. Fairhurst, An effective template protection method for face and voice cancellable identification, Int. J. Hybrid Intel. Syst. 11 (3) (2014) 157–166.
[10] A. Cavoukian, A. Stoianov, Biometric encryption: The new breed of untraceable biometrics, Biometrics: fundamentals, theory, and systems, Wiley, 2009.
[11] European Parliament, EU Regulation 2016/679 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)(2012).
[12] M. Gomez-Barrero, C. Rathgeb, et al., Protected facial biometric templates based on local gabor patterns and adaptive bloom filters, in: Proceedings of ICPR, 2014, pp. 4483–4488.
[13] J. Gonzalez-Rodriguez, J. Fierrez, et al., Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems, Forensic Sci. Int. 155 (2) (2005) 126–140.
[14] M. Günther, R. Wallace, S. Marcel, An open source framework for standardized comparisons of face recognition algorithms, in: Proceedings of ECCV, in: LNCS, 7585, 2012, pp. 547–556.
[15] J. Hermans, B. Mennink, R. Peeters, When a bloom filter is a doom filter: Security assessment of a novel iris biometric, in: Proceedings of BIOSIG, 2014.
[16] ISO/IEC JTC1 SC27 Security Techniques, ISO/IEC 24745:2011. Information Technology - Security Techniques - Biometric Information Protection, ISO, 2011.
[17] ISO/IEC TC JTC1 SC37 Biometrics, ISO/IEC 19795-1:2006. Information Technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework, ISO and IEC, 2006.
[18] A.K. Jain, Technology: Biometric recognition, Nature 449 (2007) 38–49.
[19] A. Jegede, N.I. Udzir, A. Abdullah, R. Mahmod, Face recognition and template protection with shielding function (2015).
[20] A. Kerckhoffs, La cryptographie militaire, Journal des sciences militaires 9 (1883) 5–83.
[21] Y. Kim, K. Toh, A method to enhance face biometric security, in: Proceedings of BTAS, 2007, pp. 1–6.
[22] S. Kullback, R.A. Leibler, On information and sufficiency, Annals Math. Stat. 22 (1) (1951) 79–86.
[23] N. Lalithamani, D. Sabrigiriraj, Technique to generate a face and palm vein-based fuzzy vault for a multi-biometric cryptosystem, Mach. Graph. Vis. 23 (1/2) (2014) 97–114.
[24] G. Li, B. Yang, C. Rathgeb, C. Busch, Towards generating protected fingerprint templates based on bloom filters, in: Proceedings of International Workshop on Biometrics and Forensics (IWBF), 2015.
[25] K. Nandakumar, A.K. Jain, Biometric template protection: Bridging the performance gap between theory and practice, IEEE Sig. Process. Mag. Special Issue Biometric Security Privacy (2015) 1–12.
[26] D.C.L. Ngo, A.B.J. Teoh, J. Hu, Biometric Security, Cambridge Scholars Publishing, 2015.
[27] J. Ortega-Garcia, J. Fierrez, et al., The multi-scenario multi-environment BioSecure multimodal database (BMDB), IEEE Trans. Pattern Anal. Mach. Intel. 32 (2010) 1097–1111.
[28] A. Othman, A. Ross, Privacy of facial soft biometrics: Suppressing gender but retaining identity, in: European Conference on Computer Vision Workshops, ECCV, 2014, pp. 682–696.
[29] V.M. Patel, N. Ratha, R. Chellappa, Cancelable biometrics: A review, IEEE Sig. Process. Mag. 32 (5) (2015) 54–65.
[30] P.P. Paul, M. Gavrilova, Multimodal biometric approach for cancelable face template generation, SPIE Defense, Security, and Sensing, 2012. 84070H–84070H
[31] S. Rane, Standardization of biometric template protection, IEEE Multimedia 21 (4) (2014) 94–99.
[32] N. Ratha, J. Connell, R. Bolle, Enhancing security and privacy in biometrics-based authentication systems, IBM Syst. J. 40 (3) (2001) 614–634, doi:10.1147/sj.403.0614.
[33] C. Rathgeb, F. Breitinger, C. Busch, Alignment-free cancelable iris biometric templates based on adaptive bloom filters, in: Proceedings of ICB, 2013, pp. 1–8.
[34] C. Rathgeb, M. Gomez-Barrero, et al., Towards cancelable multi-biometrics based on bloom filters: A case study on feature level fusion of face and iris, in: Proceedings of IWBF, 2015, pp. 1–7.
[35] C. Rathgeb, A. Uhl, A survey on biometric cryptosystems and cancelable biometrics, EURASIP J. Inf. Security 2011 (3) (2011).
[36] K. Simoens, B. Yang, et al., Criteria towards metrics for benchmarking template protection algorithms, in: Proceedings of ICB, 2012, pp. 498–505.
[37] A.B.J. Teoh, Y.W. Kuan, S. Lee, Cancellable biometrics and annotations on biohash, Pattern Recognit. 41 (6) (2008) 2034–2044.
[38] A.B.J. Teoh, D.C.L. Ngo, A. Goh, Personalised cryptographic key generation based on FaceHashing, Comput. Security (23) (2004) 606–614.
[39] S. Wang, J. Hu, Design of alignment-free cancelable fingerprint templates via curtailed circular convolution, Pattern Recognit. 47 (3) (2014) 1321–1329.
[40] W. Zhang, S. Shan, et al., Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition, in: Proc. ICCV, volume 1, 2005, pp. 786–791.