applied sciences

MDPI

Article

Multimodal Biometric Template Protection Based on a Cancelable SoftmaxOut Fusion Network

Jihyeon KIM [1], Yoon Gyo Jung [2] and Andrew Beng Jin Teoh [1,*]

[1] School of Electrical and Electronic Engineering, College of Engineering, Yonsei University, Seoul 03722, Korea; kim_jihyeon@yonsei.ac.kr
[2] Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA; jungyg92@gmail.com
* Correspondence: bjteoh@yonsei.ac.kr

Abstract: Authentication systems that employ biometrics are commonplace, as they offer a convenient means of authenticating an individual's identity. However, these systems give rise to concerns about security and privacy due to insecure template management. As a remedy, biometric template protection (BTP) has been developed. Cancelable biometrics is a non-invertible form of BTP in which the templates are changeable. This paper proposes a deep-learning-based end-to-end multimodal cancelable biometrics scheme called cancelable SoftmaxOut fusion network (CSMoFN). By end-to-end, we mean a model that receives raw biometric data as input and produces a protected template as output. CSMoFN combines two biometric traits, the face and the periocular region, and is composed of three modules: a feature extraction and fusion module, a permutation SoftmaxOut transformation module, and a multiplication-diagonal compression module. The first module carries out feature extraction and fusion, while the second and third are responsible for the hashing of fused features and compression. In addition, our network is equipped with dual template-changeability mechanisms with user-specific seeded permutation and binary random projection. CSMoFN is trained by minimizing the ArcFace loss and the pairwise angular loss. We evaluate the network, using six face–periocular multimodal datasets, in terms of its verification performance, unlinkability, revocability, and non-invertibility.

Keywords: cancelable biometrics system; multimodal biometrics; neural network model; security and privacy; fusion of features; random projection; authentication

Citation: KIM, J.; Jung, Y.G.; Teoh, A.B.J. Multimodal Biometric Template Protection Based on a Cancelable SoftmaxOut Fusion Network. *Appl. Sci.* **2022**, *12*, 2023. https://doi.org/10.3390/app12042023

Academic Editor: Cheonshik Kim

Received: 22 January 2022
Accepted: 14 February 2022
Published: 15 February 2022

1. Introduction

The scope of deployment of biometrics-based systems is rapidly expanding. In particular, the use of systems that rely on biometrics, such as mobile, banking, and online systems, is increasing. Biometrics capture unique physiological or behavioral trait information about users, and are therefore a convenient and highly accurate means of identity management. However, a biometric trait cannot be used if it has been exposed or abused even once, and the same biometric templates cannot be stored across multiple devices, which may increase the risk of a cross-matching attack [1,2]. A system employing biometrics must therefore prioritize security, privacy, and accuracy when authenticating individuals.

Cancelable biometrics (CB), a biometric template protection method, has been proposed to address the abovementioned concerns. A CB scheme generally consists of a feature extractor, a user-specific parameterized transformation function, and a matcher, as shown in Figure 1. CB does not directly use the original biometric template for matching, but instead uses the results of a non-invertible transformation. The original biometric data cannot be restored after being transformed, although a CB template can be generated immediately from its original counterpart. Furthermore, cross-matching of different CB templates generated from the same biometric template is highly unlikely. In brief, there are four conditions that need to be satisfied by the CB scheme, as follows [3]:

- *Non-invertibility*: It must be extremely difficult to restore the original biometric template from the CB template.
- *Revocability*: If a CB template is exposed, a new template should be generated immediately from the original biometric data. This implies that there should be no limit on the number of CB templates generated from one biometric template.
- *Unlinkability*: Two or more CB templates generated by the same user should not be distinguishable, in order to reduce the risk of a cross-matching attack.
- *Performance*: The accuracy performance of a CB-based system should not be poorer than its original counterpart.

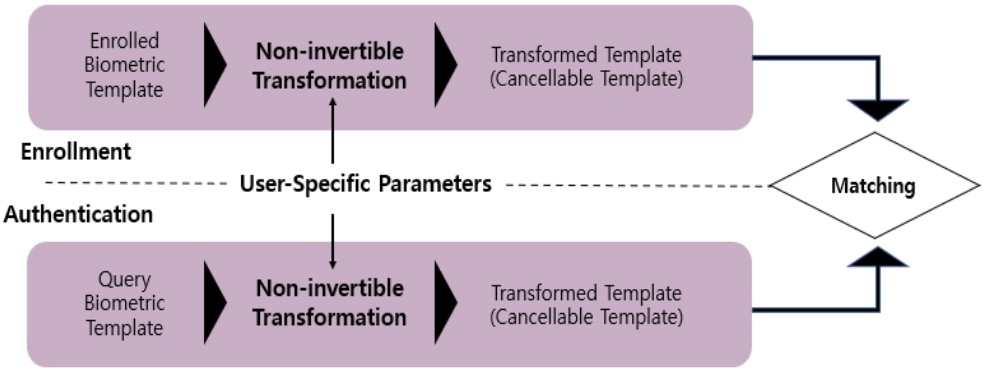


Figure 1. Illustration of a general cancelable biometrics scheme.

In general, biometric systems can be classified into two types based on the number of biometric modalities used, namely unimodal and multimodal biometrics [4,5]. A unimodal system recognizes a user based on a single biometric modality, whereas a multimodal system performs recognition based on more than one biometric modality. Unimodal biometrics has traditionally been applied, and although its performance has been proven, it has certain limitations. Since only a single biometric modality is deployed, the presence of sensor noise may affect the accuracy performance. Other problems may also arise, such as non-universality, vulnerabilities to spoofing attacks, and intra-class and inter-class similarities [6]. Multimodal biometrics is an approach that can compensate for these limitations. Since multimodal biometrics uses multiple biometric modalities, the probability of modalities being unavailable or missing is low. The accuracy performance can also be improved due to the fusion of biometrics information. Furthermore, the use of multiple modalities increases the robustness to security attacks such as spoofing [7]. However, the risk of template abuse and attack remains the same as for unimodal biometric systems, and the consequences could be catastrophic, as more private information about the user would be revealed from multiple compromised templates.

In this paper, we propose a deep-learning-based end-to-end multimodal CB scheme called cancelable SoftmaxOut fusion network (CSMoFN). By end-to-end, we mean a model that receives raw biometric data as input and produces a CB template as output [8]. Our model relies on two biometric traits, the face and periocular region, as input and produces a CB template as output.

1.1. Related Work

1.1.1. Multimodal Biometrics with Deep Learning

In this subsection, we review several works that have applied deep learning to multimodal biometric systems. Ding et al. [9] proposed a multimodal face recognition system composed of global facial features, rendered a frontal face image using a 3D face model, and uniformly sampled local face image patches. A combination of multiple convolution neural networks (CNNs) and a stacked autoencoder were used for feature learning and performing feature-level fusion.

Al-Waisy et al. [10] outlined a multimodal biometric system known as IrisConvNet. The IrisConvNet fuses left and right irises at the ranking-level. Alay et al. [11] considered iris, face, and finger veins and processed them with separate CNNs, with the output of each network fused at the score level. Gunasekaran et al. [12] also proposed a deep multimodal biometric system consisting of the iris, face, and fingerprint, called deep contourlet derivative weighted rank (DCD-WR) network. The matching is achieved with a deep learning template matching algorithm.

The study in [13] presented a multifeature deep learning network (MDLN) architecture that fused the facial and periocular regions, with the addition of texture descriptors. MDLN was designed as a feature-level fusion approach that correlated raw biometric data with texture descriptors to produce a new representation.

Algashaam [14] fused the iris and periocular region using a hierarchical fusion network. Their network allowed the system to automatically explore and discover the best strategy for combining the individual biometric scores. Luo et al. [15] outlined a deep neural network for fusion of the iris and periocular features. A co-attention feature fusion module was used to fuse the features adaptively, to obtain iris-periocular features for accurate recognition.

Jung et al. [16] proposed a teacher–student network for periocular representation learning. The teacher network, which is pre-trained with face images, is leveraged to regulate the student (periocular) network in order to enhance the periocular performance. Soleymani et al. [17] suggested a generalized compact bilinear fusion algorithm composed of multiple CNNs. Three multimodal features (iris, face, and fingerprint) are fused through a fully connected layer.

In summary, deep learning is a natural and well-suited approach for multimodal biometric systems, as deep neural networks enable feature extraction, fusion, and authentication to be performed "under one roof". However, this approach does not consider the issue of template protection.

1.1.2. Cancelable Multimodal Biometrics with Deep Learning

Multimodal biometrics with template protection is not a new topic, with many papers having been published on this subject [18–22]. However, deep-learning-based multimodal biometric systems that include template protection remain very scarce.

Abdellatef et al. [23] designed a multi-instance CB for the face using multiple CNNs to extract features from multiple regions of a face image, such as face, eyes, nose, mouth, etc. After fusion of several deep features, a cancelable template was generated via bio-convolving encryption. Their method achieved a better performance than when a unimodal biometric was applied. However, a detailed analysis of CB design criteria was not provided.

Talreja et al. [24] introduced a method for generating CB templates from face and iris biometrics. Each biometric image was subjected to feature extraction via a CNN, with a random component selected from the generated features and used as a transformation key. The transformed templates generated from the CB module were then converted to a secure sketch via a forward error correction (FEC) decoder and cryptographic hashing. However, a compromised transformation key may pose the risk of CB template inversion.

Sudhakar et al. [25] put forward a finger vein and iris-based CB scheme in which a CNN was applied to perform feature extraction and a support vector machine (SVM) was used for user verification. The template was protected with a random projection-based approach.

More recently, El-Rahiem et al. [26] proposed a multibiometric CB system using fingerprint, finger vein, and iris images. First, feature extraction was performed for each biometric modality through a CNN and fusion of the three modalities was achieved using feature maps. The DeepDream algorithm was then applied to give a cancelable template. However, a security analysis was not performed.

A detailed comparison with the above papers is provided in Section 4.6.

1.2. Motivations and Contributions

In this paper, we propose an end-to-end CSMoFN scheme for multimodal biometric systems. Our network fuses face and periocular biometric traits at the feature level, producing a single CB template as output. More specifically, CSMoFN is composed of two components, the first of which transforms the fused biometric vectors based on the notion of random permutation maxout transform (RPMoT), which was proposed in [27]. Although RPMoT is a CB transformation scheme, it is not learnable and is data-agnostic, and hence barely meets the performance requirements for the CB scheme. RPMoT is reformulated as a part of the deep neural network and is referred to here as the permutation SoftmaxOut transform (PSMoT). PSMoT is data-driven, learnable, and parameterized by user-specific permutation seeds to satisfy the revocability and unlinkability criteria. PSMoT can be viewed as a locality-sensitive hashing process that transforms biometric data from a high-dimensional input space to a relatively low-dimensional hash space [28].

The PSMoT comes with a customized layer called permutation SoftmaxOut layer. The layer composed of maxout units and a modified softmax function to approximate the permutation and the winner-takes-all operations in the RPMoT. Apart from that, the modified softmax function is also useful to minimize the quantization errors introduced by the SoftmaxOut approximation. Since PSMoT aims to produce a discrete hash vector from the network directly, it is a representation learning problem. Hence, a pairwise distance-based loss called Pairwise Angular (PA) loss, is introduced to optimize the margin between intra- and inter-class distances.

The output of PSMoT (i.e., a hash vector) is immediately followed by the second component of the network, called the multiplication-diagonal compression (MDC) module. The MDC module is designed to further enhance the security and compress the PSMoT hash vector, and offers a user-specific seeded binary random projection mechanism as a means to enhance the revocability and unlinkability of the proposed method.

Both the PSMoT and MDC transformation follow the many-to-one mapping principle attributed to their hashing trait, meaning that inversion (one-to-many mapping) of the terminal hash output is theoretically impossible and computationally hard in practice. This is essential to satisfy the non-invertibility requirement for a CB scheme.

In this paper, we opt to fuse the face and periocular features. The periocular region, also known as the periphery of the ocular area, includes the vicinity of a person's eyes and contains information on the subject's eyebrows, eyelashes, and skin texture. The periocular region is a complementary biometric of the face that is helpful in terms of enhancing the biometric performance of the face alone. It is particularly useful in situations such as when a mask is worn, where the face is occluded and a performance degradation is expected. As we will show in the experiment section, fusion of the face and periocular region outperforms the respective unimodal counterparts, i.e., the face or periocular region alone.

We can summarize our contributions as follows:

- A deep-learning-based CB scheme for multimodal biometrics is proposed. Although the face and periocular biometrics form the focus of this paper, our proposed method can also be applied to other biometrics modalities, provided the input is a raw image.
- A deep network, CSMoFN is composed of three modules: a feature extraction and fusion module, a PSMoT module, and an MDC module is proposed to realize the above proposal. The first module is responsible for performing feature extraction and fusion and the latter two are cancelable transformation functions, which are devised with respect to the four CB design criteria.
- The three modules are trained in an end-to-end manner with a combination of classification loss and representation learning, namely ArcFace loss and PA loss.
- We evaluate the proposed network on six face–periocular multimodal datasets in terms of verification performance, unlinkability, revocability, and non-invertibility.

2. Preliminaries: Random Permutation Maxout Transform

RPMoT [27] is a data-agnostic CB scheme that transforms a biometric feature vector into a discrete hash vector, as illustrated in Figure 2. RPMoT is parameterized by a user-specific seeded permutation matrix, which means that the hash vector can be revoked if it is compromised. The flow of the algorithm is summarized as follows:

1. A user-specific permutation matrix is first created. Suppose the size of the biometric feature vector $\boldsymbol{X}$ is d and permutation matrix is $d \times d$. There are m permutation matrices that are generated and stacked to form $\boldsymbol{P}$.
2. $\boldsymbol{X}$ and $\boldsymbol{P}$ are multiplied to yield a matrix $\boldsymbol{W}$ with size $m \times d$.
3. The first q column vectors of $\boldsymbol{W}$ are used and the rest are discarded, yielding a matrix $\boldsymbol{Y}$ with size $m \times q$.
4. Finally, the position of the feature with the largest value in each row of $\boldsymbol{Y}$ is recorded as the index value. When all rows have been processed, the RPMoT hash vector $\boldsymbol{u}$ with size m is obtained. Note that $\boldsymbol{u}$ is an integer-value vector ranging from 1 to q.

In this paper, RPMoT is redesigned as a component of CSMoFN, which is learnable and data-driven. However, the essence of PRMoT as a CB scheme that satisfies the requirements of non-invertibility, revocability, and unlinkability remains intact.

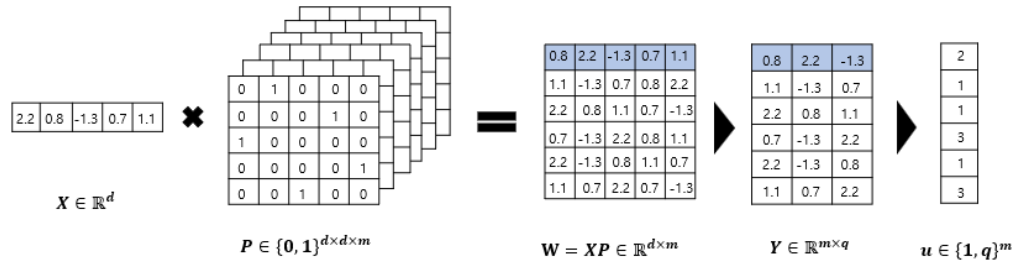


Figure 2. Illustration of RPMoT, with $d = 5$, $m = 6$, and, $q = 3$.

3. Proposed Method

3.1. Overview

The proposed CSMoFN system takes face and periocular biometric information as its input and is composed of three modules: a feature extraction and fusion module, a PSMoT module, and an MDC module. As portrayed in Figure 3, the backbone of CSMoFN is based on a CNN, which performs feature extraction from images of faces and periocular regions via multiple convolutional blocks. The extracted features are fused at the feature level. PSMoT then transforms the fused vector to a discrete hash vector, which is further compressed to yield a terminal hash vector (CB template) from the MDC module. A user-specific token or password is required to generate a random seed for permutation and binary random projection in the PSMoT and MDC modules, respectively.

In essence, the proposed system is a two-factor cancelable multimodal biometric system for which both biometric inputs and user-specific token/passwords are required. The entire network is trained end-to-end following an open-set (database and identity independence) evaluation protocol [29]. This means that the model is trained on datasets that are independent of the enrolled subjects, which is preferable for biometric systems as the model does not need to be retrained when a new user is enrolled or an old CB template is reissued. In the latter case, the user only needs to change the token or password.

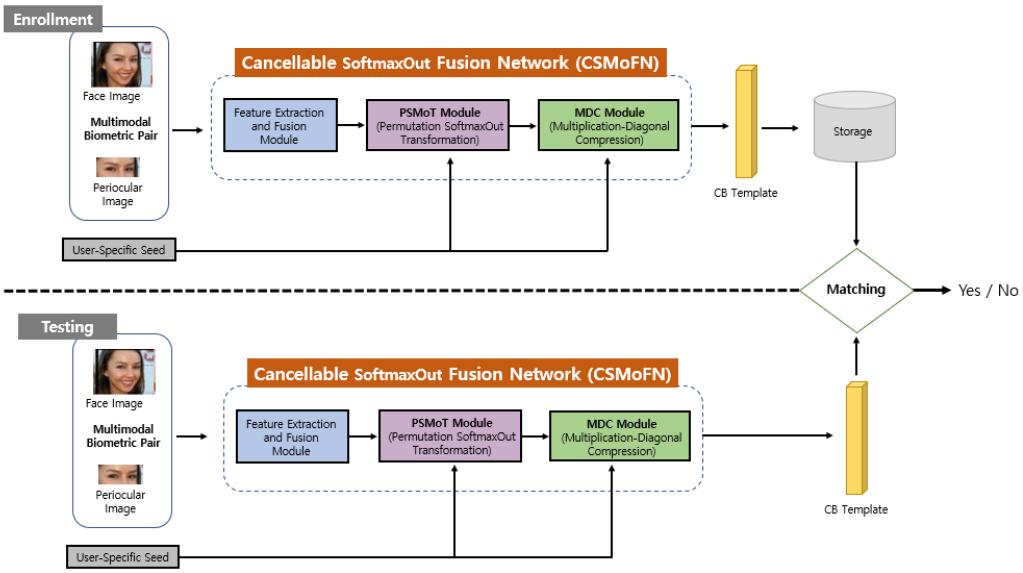


Figure 3. Overview of the proposed CSMoFN model.

3.2. Feature Extraction and Fusion Module

In Figure 4, we adopt ResNet-50 as the backbone for the proposed method, which consists of 49 convolution layers and a linear activated fully connected layer with p neurons, thus producing a p-dimensional feature vector for each face and periocular image. The backbone is pre-trained using the MS-Celeb-1M dataset [30]. Two feature vectors from the face $\mathbf{z}_{face}$ and periocular region $\mathbf{z}_{periocular}$ are fused at the feature level by concatenation, and hence the number of dimensions of the fused vector $\mathbf{z} = [\mathbf{z}_{face}\ \mathbf{z}_{periocular}]$ is $2p$. Fusion with concatenation can largely preserve the useful information from both biometrics despite the increase in feature size compared to other strategies such as the feature sum or average. In this work, we set $p = 512$.

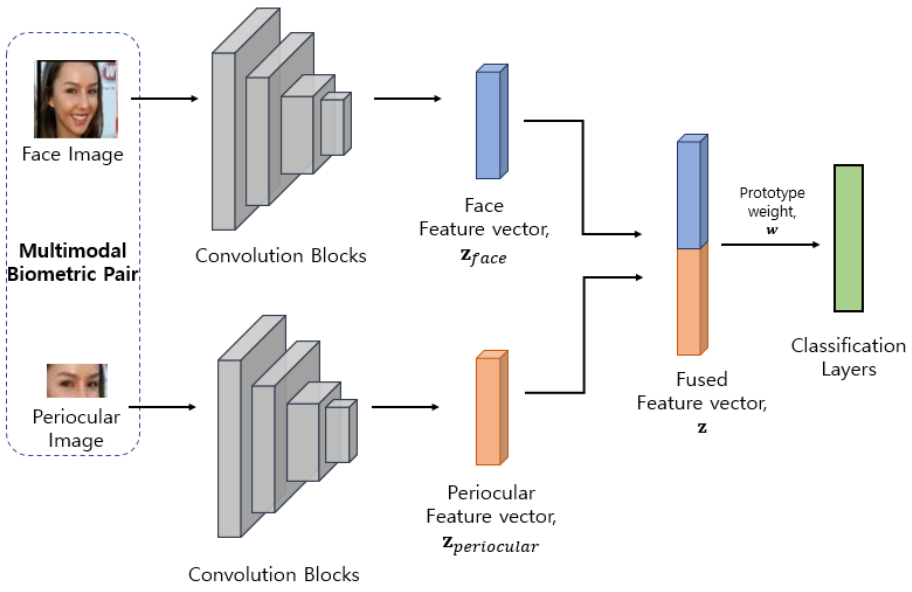


Figure 4. The feature extraction and fusion module.

3.3. Permutation SoftmaxOut Transform (PSMoT) Module

The PSMoT module is located immediately after the FC (Fully Connected) layer of the feature extraction and fusion module. As depicted in Figure 5, it is composed of two ReLU activated hidden layers, the first of which ($\boldsymbol{h}_1$) consists of l_1 neurons, and the second

(h_2) consists of l_2 neurons for nonlinear transformation purposes. We set $l_1 = l_2 = 2014$. The SoftmaxOut layer is a dedicated layer designed for hashing. There are m maxout units $\boldsymbol{m}_i$ composed of q permutable linear activated neurons. A permutation has user-specific and/or application-specific dependence. The maxout unit is a function that returns the index of the maximal entry of the q neurons. The hashing layer produces m discrete hash codes v_i forming a hash vector $\boldsymbol{v} \in \{1, \ldots, q\}^m$.

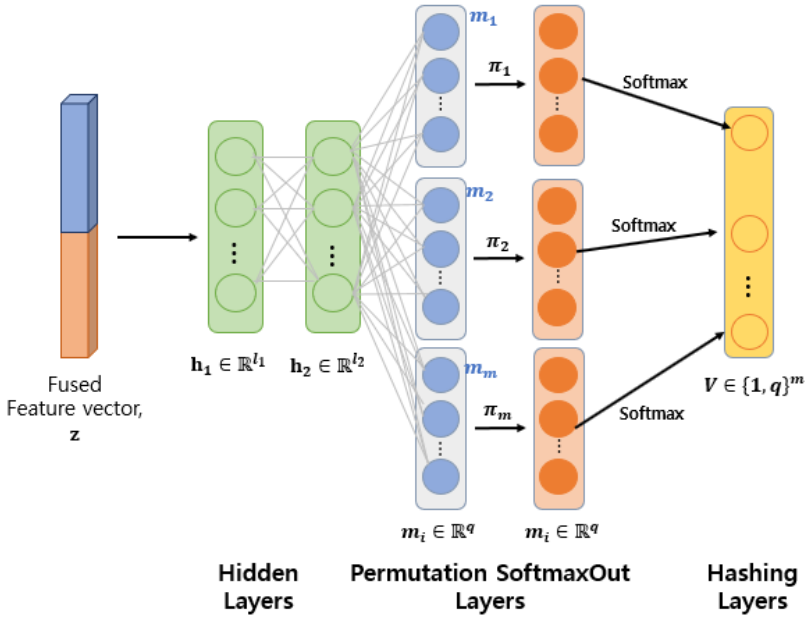


Figure 5. The permutation SoftmaxOut transform (PSMoT) module.

Recall that the RPMoT (Section 2) produces the index value of the maximum entry of a q-dimensional permuted vector, as described in Step 4. This is equivalent to taking the index value v_i from a permutable maxout unit, as follows [31]:

$$v_i = \underset{q}{\operatorname{argmax}} \pi_i(\boldsymbol{m}_i) \in \{1, \ldots, q\},\ i = 1, \ldots, m \tag{1}$$

However, Equation (1) is non-differentiable and hence non-trainable with backpropagation. In view of this, we approximate Equation (1) with the following function:

$$v_i = \sum_{j=1}^{q} j s_\mu(\pi_i(\boldsymbol{m}_i)) \in \{1, \ldots, q\} \tag{2}$$

where $s_\mu()$ is the Softmax function parameterized with $\mu > 1$:

$$s_\mu(v) = \frac{e^{\mu v}}{\sum_{i=1}^{q} e^{\mu v_i}} \tag{3}$$

Unlike the conventional Softmax function, $s_\mu(v)$ is parameterized by a scalar factor $\mu > 1$ that forces the output of the network towards zero or one, thereby allowing the PSMoT to learn a discrete hash code. In our experiments, we use $\mu = 9$.

Unlike RPMoT, which is data-agnostic, PSMoT is data-driven. In addition, RPMoT transforms a biometric feature vector that is separately processed by a feature extractor, whereas the PSMoT module is connected to the CNN backbone and both are trained in an end-to-end manner. The inclusion of two hidden layers is beneficial in terms of improving the feature discrimination, which can be attributed to the nonlinear transformation of the fused features.

3.4. Multiplication-Diagonal Compression (MDC) Module

MDC is a learning-free module located immediately after the PSMoT. As shown in Figure 6, the PSMoT hash vector with size m is first reshaped into a matrix $\boldsymbol{V}'$ with size $k \times n$, where $k = m/n$. Then, based on a user-specific seed, a binary random matrix $\boldsymbol{R} \in \{0, 1\}^{n \times k}$ is generated and multiplied with $\boldsymbol{V}'$, yielding a matrix $\boldsymbol{Q} \in \{0, 1\}^{k \times k}$. Finally, the diagonal elements of $\boldsymbol{Q}$ are extracted and a terminal hash vector (CB template) $\boldsymbol{s} \in \{1, \ldots, q\}^k$ is obtained.

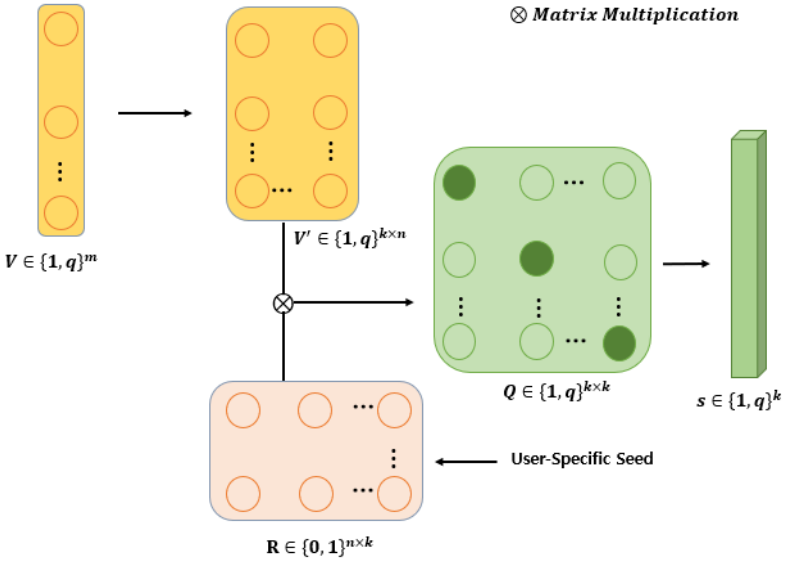


Figure 6. Structure of the multiplication-diagonal compression module.

The MDC module is devised to further enhance the non-invertibility, revocability, and unlinkability of the CSMoFN. Another important goal for the MDC module is to compress the PSMoT hash vector from size m to k (where $k \ll m$) without sacrificing accuracy. Offering computational advantages, the multiplication of the user-specific binary random projection and the hash matrix is a special kind of random projection [32] that can approximately preserve the pairwise distances of the hash vectors with respect to the distance of their original counterpart [33].

Finally, the extraction of the diagonal elements from $\boldsymbol{Q}$ can be seen as yet another many-to-one mapping that enhances the non-invertibility property of the proposed scheme.

3.5. Loss Function

- **ArcFace Loss**

The ArcFace loss [34] is used in the feature extraction and fusion module as a means of enhancing the terminal hash code discrimination. It is a modified Softmax classification loss in which the prototype weight vector of the j^{th} identity $\mathbf{w}_j$ is L2-normalized. Specifically, the target logit (activation of the classification layer before applying the Softmax function) is redefined as $\mathbf{w}_j^T \mathbf{z}_i = ||\mathbf{w}_j|| ||\mathbf{z}_i|| \cos\theta_j$, where $\mathbf{z}_i$ are the L2-normalized fused features (Section 3.2) of the i^{th} sample, belonging to the j^{th} identity. The normalization of the fused features and weights means that the predictions rely only on the angle between $\mathbf{z}_i$ and $\mathbf{w}_j$, denoted as θ_j. The prediction of the y_ith identity then only depends on θ_{y_i}. The ArcFace loss is defined as:

$$\mathcal{L}_{ArcFace} = -\frac{1}{B} \sum_{i=1}^{B} \log \frac{e^{\gamma(\cos(\theta_{y_i} + \beta))}}{e^{\gamma(\cos(\theta_{y_i} + \beta))} + \sum_{j=1, j \neq y_i}^{N} e^{\gamma \cos\theta_j}} \tag{4}$$

where B is the batch size, β is an angular margin introduced to force the classification boundary closer to that prototype weight $\mathbf{w}_j$, and γ is a feature rescaling factor. In this

manner, the learned fused features are distributed on a hypersphere with a radius of γ. In this paper, we set $\beta = 0.35$ and $\gamma = 25$.

- **Pairwise Angular Loss**

The pairwise angular (PA) loss function is introduced to optimize the PSMoT module. Face–periocular pairs are associated with similarity labels c_{ij}, where c_{ij}= 1 implies that two face–periocular pairs are from the same identity (positive pair) and c_{ij}= 0 indicates that they are from different identities (negative pair). The aim of the loss function is to ensure that the similarity between a pair of hash vectors $\mathbf{s}_i$ and $\mathbf{s}_j$ in the MDC module is high if they are positive pairs, and to make the dissimilarity greater than a given margin for negative pairs. Our PA loss is defined as follows:

$$\begin{aligned} \mathcal{L}_{PA} = \quad & -log\left[c_{ij}\frac{e^{\delta\cos(\phi+\alpha)}}{e^{\delta\cos(\phi+\alpha)}+e^{\delta\sin\phi}}\right] \\ & -log\left[(1-c_{ij})\frac{e^{\delta\sin(\phi+\alpha)}}{e^{\delta\sin(\phi+\alpha)}+e^{\delta\cos\phi}}\right] \end{aligned} \tag{5}$$

where δ is a scaling factor, α is the angular margin, and ϕ is the cosine distance between $\hat{s}_i$ and $\hat{s}_j$ (i.e., $\phi = \cos^{-1}\left(\hat{s}_i^T\hat{s}_j\right)$, given that $\hat{s}$ is the L2-normalized hash code to be rescaled based on δ).

Similarly, to the ArcFace loss, normalization and rescaling on $\hat{s}_i$ and $\hat{s}_j$ means that the similarity measure relies only on the angle between the two hash codes, and thus forces the hash codes to be set down on a hypersphere with radius δ. An additive angular margin penalty α between $\hat{s}_i$ and $\hat{s}_j$ is introduced to enhance the intra-class compactness and the inter-class separation simultaneously. Here, we set $\delta = 2.5$ and α = 0.5.

- **Total Loss**

In a nutshell, the total loss function $\mathcal{L}_{total}$ used to optimize the CSMoFN is given as:

$$\min\mathcal{L}_{total} = \mathcal{L}_{ArcFace} + \mathcal{L}_{PA} + \alpha L_2 \tag{6}$$

where L_2 is a weight decay regularizer and α is a coefficient that is beneficial to reduce overfitting. In this paper, we set $\alpha = 0.5$.

4. Experiments

4.1. Datasets

Our experiments were performed on six face–periocular datasets. Although these datasets originally contained only face images, periocular images were obtained by cropping the eye region from the face images. The six datasets considered are AR [35], Ethnic [36], Facescrub [37], IMDB Wiki [38], Pubfig [39], and YTF [40].

- **AR** was generated by the Computer Vision Center (CVC) at Universitat Autonoma de Barcelona and consists of over 4000 frontal view color images of 126 subjects. It was constructed under strictly controlled conditions, and each image shows a different facial expression, with different illumination conditions and occlusions.
- **Ethnic** is a large collection dataset composed of subjects of different ethnicities. All periocular images were obtained under various uncontrolled conditions as found in a wild environment (i.e., with variations in camera distance, pose, and location). It consists of 85,394 images of 1034 subjects.
- **Facescrub** is a large face dataset composed of 530 celebrity face images. It consists of about 200 images per person, with a total of 106,863 images. Images were retrieved from the Internet and shot in a real-world environment, i.e., under uncontrolled conditions.
- **IMDB Wiki** is a large dataset of celebrities, including their birthdays, names, genders, and related images. A dataset was constructed by obtaining meta-information from Wikipedia as well as the IMDB website. It consists of a total of 523,051 face images

of 20,284 celebrities, of which 460,723 images were drawn from IMDB and 62,328 from Wikipedia.

- **Pubfig** (Public Figures Face) consists of 58,797 images of 200 people obtained from the Internet. Images were taken in completely uncontrolled situations and with non-cooperative subjects. It therefore has large variations in the characteristics of the environment, such as pose, lighting, expression, and camera.
- **YTF** (YouTube Face) dataset consists of 1595 subjects and 3425 videos. All videos were downloaded from YouTube, with an average of 2.15 videos per subject, and each video ranged from 48 to 6070 frames.

We followed the open-set evaluation protocol [29], in which the training and testing datasets do not overlap. The training set was constructed from the Ethnic and IMDB Wiki subsets, and the subjects were independent of the testing sets. Six datasets were used for testing. Table 1 gives a summary of the composition of each of the training and testing datasets.

Table 1. Description of training and testing datasets.

	Training Set	**Testing Set**					
		AR	**Ethnic**	**Facescrub**	**IMDB Wiki**	**Pubfig**	**YTF**
No. of subjects	1054	126	325	530	2129	200	1595
No. of images	166,737	700	1645	31,066	40,241	9220	150,259

4.2. Experimental Setup

Evaluations were carried out under authentication (verification). The equal error rate (EER) and the receiver operating characteristic (ROC) curve were used as authentication metrics for the proposed method. The specifications of the computer used for the experiment were an Intel(R) Core (TM) i7-6700 K CPU @ 4.00 GHz, 32 GB of RAM, and NVIDIA GeForce GTX 1080 Ti GPU, with the model implemented using the Pytorch library.

For network training, the batch size B was fixed to 256, the epoch was 90, and the learning rate was 0.0001. Matching of the hash vectors was performed using the Hamming distance. All experiments were conducted using the same user-specific seeds and a scenario called the stolen-token scenario [27], to enable a fair comparison.

4.3. Hyperparameter Analysis

In this section, we explore the impact of the two essential hyperparameters, q and m, used in CSMoFN, where q determines the range of elements of the PSMoT hash vector and m is the hash vector size. We set $m = 256,\ 512,\ 1024,\ 2048,$ and 4096, and $q = 8,\ 16,\ 32,$ and 64.

Table 2 shows the performance on the six datasets at various settings of m and a fixed value of q = 32. For all datasets, it can be seen that the smaller the value of m, the lower the accuracy performance, which implies a loss of information. As m increases, the degree of information loss decreases, which leads to a lower EER.

For Table 3, we set m = 4096 and checked the performance for different values of q. We can observe that although the parameter q does not have a significant effect on the overall accuracy performance, medium (i.e., $q = 32$) and small values (such as $q = 8$) give a slightly higher EER. However, it is better to set q to a larger value to enhance the security, as a larger q increases the complexity of a brute-force guessing attack on a hash vector.

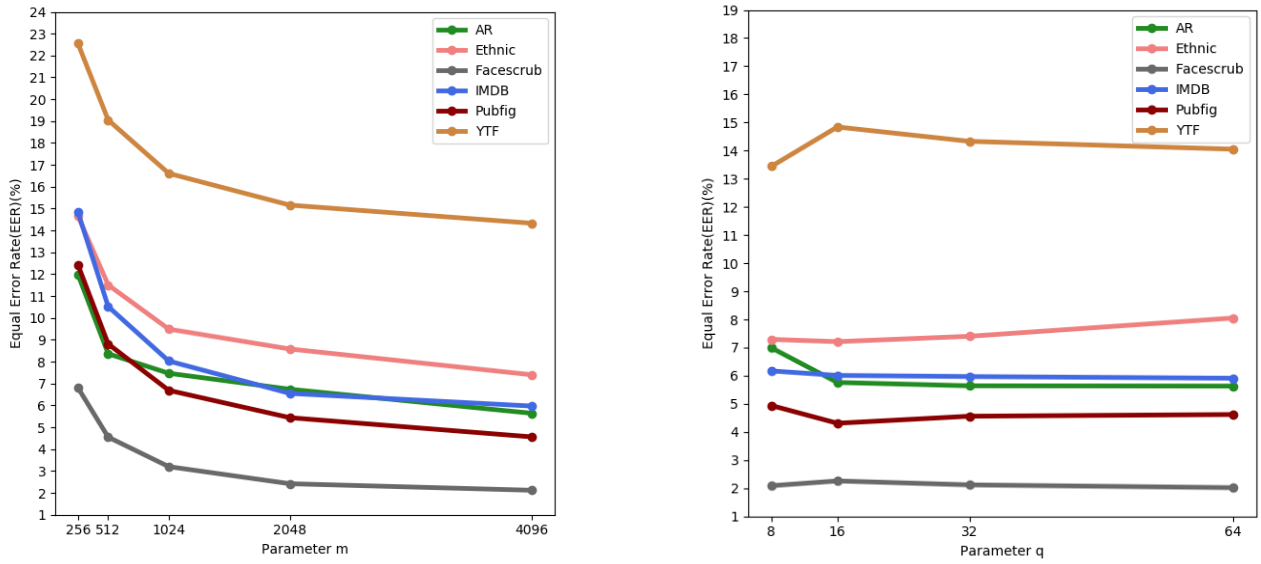
Figure 7 shows the EER (%) performance for m and q, while Figure 8 shows the ROC curve and area under the curve (AUC) for six datasets with m = 4096 and q = 32.

Table 2. Accuracy of the model for various values of m and $q = 32$.

Equal Error Rate (EER) (%)					
$q = 32$	m				
	256	**512**	**1024**	**2048**	**4096**
AR	11.99	8.36	7.47	6.74	5.64
Ethnic	14.66	11.51	9.49	8.58	7.40
Facescrub	6.81	4.55	3.20	2.42	2.12
IMDB Wiki	14.84	10.52	8.03	6.55	5.97
Pubfig	12.43	8.81	6.69	5.44	4.56
YTF	22.55	19.05	16.61	15.16	14.33
Average	13.88	10.47	8.58	7.48	6.67

Table 3. Accuracy of the model for various values of q and $m = 4096$.

Equal Error Rate (EER) (%)				
$m = 4096$	q			
	8	**16**	**32**	**64**
AR	6.99	5.76	5.64	5.63
Ethnic	7.29	7.21	7.40	8.05
Facescrub	2.09	2.26	2.12	2.02
IMDB Wiki	6.17	6.01	5.97	5.91
Pubfig	4.94	4.31	4.56	4.62
YTF	14.44	14.84	14.33	14.05
Average	6.99	6.73	**6.67**	6.71



(**a**) (**b**)

Figure 7. EER (%) for (**a**) varying m with $q = 32$ and (**b**) varying q with $m = 4096$.

As discussed in Section 3.4, the MDC module reshapes the PSMoT hash vector with size m to a matrix $\boldsymbol{V}'$ with size $k \times n$, and hence $m = k \times n$. Note that the size of the final hash vector (CB template) is k, with the value of k dependent on m and n. For this experiment, we examine two combinations of k and n, i.e., $(k, n) = (128, 16)$ and $(256, 8)$ with a fixed value of $m = 2048$. The EER is shown in Table 4. We observe that $(256, 8)$ outperforms $(128, 16)$, which implies that $k = 256$ is a better choice than $k = 128$. It is worth noting that the EERs for the PSMoT hash vector (without reshaping) and (256×8) are

identical, although the size of the final hash vector in the latter case is only 256. This demonstrates the performance of the MDC module in terms of compression.

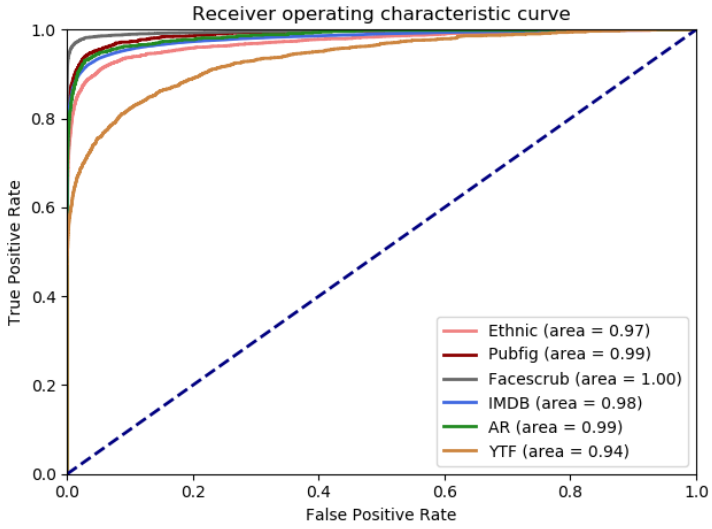


Figure 8. ROC curves and AUC analysis for six datasets with $m = 4096$ and $q = 32$.

Table 4. Performance comparison for varying values of (k, n) with $m = 2048$ and $q = 32$.

Equal Error Rate (EER) (%)			
$q = 32$	$m = 2048$		
	Without Reshaping	**(128, 16)**	**(256, 8)**
AR	6.74	7.12	6.74
Ethnic	8.58	8.88	8.58
Facescrub	2.42	3.11	2.42
IMDB Wiki	6.55	7.83	6.55
Pubfig	5.44	6.19	5.44
YTF	15.16	16.33	15.16
Average	7.48	8.24	**7.48**

4.4. Performance Comparison with Unimodal CB Systems

In this section, we demonstrate the advantage of the proposed multimodal CB system through a comparison with unimodal CB systems (i.e., where either the face or the periocular region alone is adopted).

Table 5 shows the average EER performance for two unimodal CB systems and a multimodal CB biometric system for six datasets. Figure 9 shows the change in performance for varying m and with q fixed at 32. In general, the performance of each system improves with large m and a moderate value of q. This is consistent with the finding in Section 4.3. However, we also note that the best EER can be achieved by the face–periocular CB system over its unimodal counterparts, with an EER reduction of around 50% for the unimodal systems with $m = 4096$ and $q = 32$. This suggests that fusion is essential for performance gain, despite the simplicity.

Table 5. Average performance of unimodal and multimodal CB systems on six datasets for varying values of m and q.

Equal Error Rate (EER) (%)					
Face					
q	$m(k)$				
	256(32)	**512(64)**	**1024(128)**	**2048(256)**	**4096(512)**
8	15.68	14.82	13.83	12.87	12.56
16	18.06	15.27	14.30	13.66	13.01
32	17.65	15.85	14.14	13.62	13.16
64	17.83	15.87	14.90	13.70	13.31
Periocular					
q	$m(k)$				
	256(32)	**512(64)**	**1024(128)**	**2048(256)**	**4096(512)**
8	16.35	15.43	15.61	15.00	15.20
16	18.35	17.96	16.04	15.57	15.02
32	19.44	17.75	16.74	15.94	15.33
64	19.70	18.34	17.27	15.91	15.29
Fused					
q	$m(k)$				
	256(32)	**512(64)**	**1024(128)**	**2048(256)**	**4096(512)**
8	13.63	10.73	8.53	7.35	6.99
16	14.33	10.69	8.53	7.43	6.73
32	13.88	10.47	8.58	7.48	6.67
64	13.64	10.53	8.44	7.49	6.71

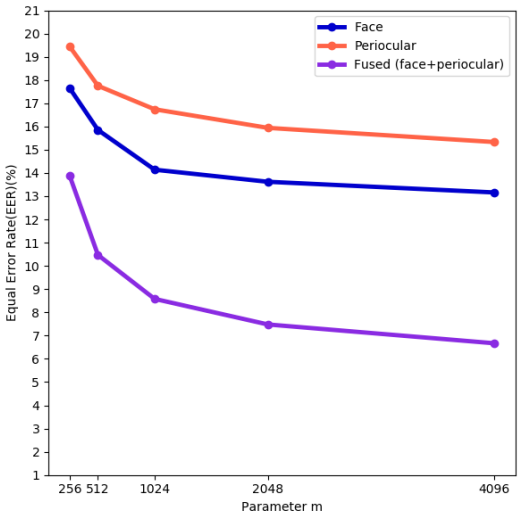


Figure 9. Average EER (%) for face, periocular, and fused biometric traits, for various m and q = 32.

4.5. Ablation Studies

This section presents an ablation study on the proposed CSMoFN. We first explore the accuracy performance of the sole feature extraction module with cosine distance, which is equivalent to an unprotected biometric system, and serves as a baseline. We then examine the feature extraction module + PSMoT, and lastly the entire CSMoFN. The latter two are CB systems.

From Table 6, we can observe that the baselines for the face and periocular region alone perform better than their CB counterparts (i.e., PSMoT and CSMoFN). This is as expected, and can be attributed to the performance–security tradeoff made in CB systems,

where the performance may be degraded after the CB transformation. However, the use of face–periocular fusion largely restores the verification performance for CSMoFN.

Table 6. Ablation study for each module (baseline, PSMoT, and CSMoFN).

Equal Error Rate (EER) (%)							
Baseline (Feature Extraction without Hashing)							
	AR	**Ethnic**	**Facescrub**	**IMDB Wiki**	**Pubfig**	**YTF**	**Average**
Face	4.23	4.06	1.58	4.79	3.50	11.84	5.00
Periocular	6.77	5.61	3.13	6.53	5.48	15.16	7.11
Fused	4.67	5.29	1.75	5.12	4.07	13.56	5.74
PSMoT ($m = 4096$, $q = 32$)							
	AR	**Ethnic**	**Facescrub**	**IMDB Wiki**	**Pubfig**	**YTF**	**Average**
Face	10.71	15.98	5.89	16.92	14.93	22.66	14.51
Periocular	15.36	16.85	10.40	18.69	16.31	22.73	16.72
Fused	5.90	7.48	2.16	5.97	4.68	14.55	6.79
CSMoFN (PSMoT + MDC)							
	AR	**Ethnic**	**Facescrub**	**IMDB Wiki**	**Pubfig**	**YTF**	**Average**
Face	9.62	15.34	5.07	15.64	12.68	20.61	13.16
Periocular	11.75	16.56	9.43	17.40	15.43	21.38	15.32
Fused	5.64	7.40	2.12	5.97	4.56	14.33	6.67

4.6. Remarks on Deep-Learning-Based Multimodal Cancelable Biometrics Schemes

In this section, we present a summary with remarks in Table 7 rather than a comparison between different approaches. This is because a fair comparison between different template protection schemes is very difficult, or even impossible, due to several factors such as the choice of biometric modality, fusion method, datasets, and evaluation metrics.

Table 7. Summary of works related to cancelable multimodal biometric systems.

Ref.	**Modalities**	**Fusion**	**Remarks**
Our proposed method	Face, periocular region	Feature-level	**Methods:** End-to-end deep-learning-based cancelable biometrics scheme with three modules (feature extraction and fusion module, permutation SoftmaxOut transformation module, multiplication-diagonal compression module). **Revocability:** ✔ **Unlinkability:** ✔ **Non-invertibility:** ✔ **Accuracy performance:** EER = 2.12% (Facescrub dataset), EER = 6.67% (average over six datasets)
Abdellatef et al. [23]	Face, eye region, nose region, mouthregion	Feature-Level	**Methods:** Feature extraction is performed with multiple CNNs. After fusion of multiple deep features, BTP transformation is performed with bioconvolving encryption. **Revocability:** ✘ **Unlinkability:** ✘ **Non-invertibility:** ✘ **Accuracy performance:** Accuracy = 93.4% (PaSC dataset)
Talreja et al. [24]	Face, iris	Feature-level	**Methods:** Deep feature extraction and binarization using CNN from multimodal modalities. A random component is selected from the generated features and used as a transformation key. The transformed templates are converted to a secure sketch via an FEC decoder and cryptographic hashing. **Revocability:**✘ **Unlinkability:** ✘ **Non-invertibility:** ✘ **Accuracy performance:** Accuracy = 99.16% (stolen key scenario)

Table 7. *Cont.*

Ref.	Modalities	Fusion	Remarks
Sudhakar et al. [25]	Finger vein, right and left irises	Feature-level	**Methods:** Feature extraction is performed through a CNN and SVM is used for verification. The template is protected with a random projection-based approach. **Revocability:** ✔ **Unlinkability:**✔ **Non-invertibility:** ✔ **Accuracy performance:** EER = 0.05% (FV-USM dataset)
El-Rahiem et al. [26]	Fingerprint, finger vein, Iris	Feature-level	**Methods:** After feature extraction with a CNN, fusion is performed through the fusion layer. A cancelable template is generated through the reconstruction process by applying the DeepDream algorithm consisting of many Convnets. **Revocability:**✘ **Unlinkability:** ✘ **Non-invertibility:**✘ **Accuracy performance:** EER = 0.0032%

Analysis keys: ✔ = Explicit, ✘ = None.

5. Unlinkability and Revocability Analysis

5.1. Unlinkability Analysis

In our unlinkability analysis, we follow the protocol and method proposed in [41]. The "mated score" and "non-mated score" first have to be calculated and are defined as follows.

Mated sample scores: This is a score calculated through cross-matching of the same subject. In our case, we use the face–periocular pair $\mathbf{X}$ of the same user and different permutations and random projection seeds $\boldsymbol{r}$.

Let the mated CB template pair be $\mathbf{T}_{\mathrm{m1}} = \mathrm{CSMoFN}(\mathbf{X}_1, \boldsymbol{r}_1)$ and $\mathbf{T}_{\mathrm{m2}} = \mathrm{CSMoFN}(\mathbf{X}_1, \boldsymbol{r}_2)$. The mated-samples score can then be obtained via $s = d_H(\mathbf{T}_{\mathrm{m1}}, \mathbf{T}_{\mathrm{m2}})$, where d_H is the Hamming distance.

The mated sample distribution is denoted as $p(s|H_m)$, where H_m belongs to the relationship in which both CB templates are mated.

Non-mated sample scores: The non-mated score is calculated in a similar way but for different subjects. Using a similar notation to that given above, let the non-mated CB template pair be $\mathbf{T}_{\mathrm{nm1}} = \mathrm{CSMoFN}(\mathbf{X}_1, \boldsymbol{r}_1)$, $\mathbf{T}_{\mathrm{nm2}} = \mathrm{CSMoFN}(\mathbf{X}_2, \boldsymbol{r}_2)$. The non-mated scores are then estimated as $s = d_H(\mathbf{T}_{\mathrm{nm1}}, \mathbf{T}_{\mathrm{nm2}})$, and the distribution of the non-mated sample scores is $p(s|H_{nm})$, where H_{nm} is when both templates are non-mated.

In addition, we use two measures of unlinkability: a local and a global measure.

Local measure $D_{\leftrightarrow}(s)$: This measure represents the likelihood ratio of two score variances, $D_{\leftrightarrow}(s) = p(H_m|s) - p(H_{nm}|s) \in [0, 1]$

Global measure $D_{\overset{sys}{\leftrightarrow}}$: Unlike the local measure, this metric evaluates the unlinkability of the overall system independently of the score domain. This measure also has a range of [0, 1].

A CB scheme is judged to ideally satisfy the unlinkability criterion if $p(H_m|s) = p(H_{nm}|s)$. If they are completely separated, the CB templates are fully linkable; in other words, if both the local and global measures are close to zero, the CB scheme is deemed nonlinkable.

According to the proposed benchmark protocol in [41], we carried out experiments by generating three CSMoFN hashed vectors with m = 4096 and q = 32, by using the Pubfig, Facescrub, and YTF datasets with different user-specific seeds. The three distributions, the mated samples score, non-mated samples score, and local measure values are all plotted together in Figure 10. It can be seen that the two score distributions, mated and non-mated, explicitly overlap. Furthermore, the meaning of this demonstrates that CB templates are unlinkable. Furthermore, the global measure $D_{\overset{sys}{\leftrightarrow}}$ of three datasets are 0.056, 0.039, and 0.104, respectively. For each specific linkage score s, $D_{\leftrightarrow}(s) = 0$ denotes fully unlinkability, while $D_{\leftrightarrow}(s) = 1$ is a fully linkable of two transformed templates. With the significant

overlap, the overall linkability of the proposed method is close to zero. This indicates that the CSMoFN hashed vectors are unlinkable.

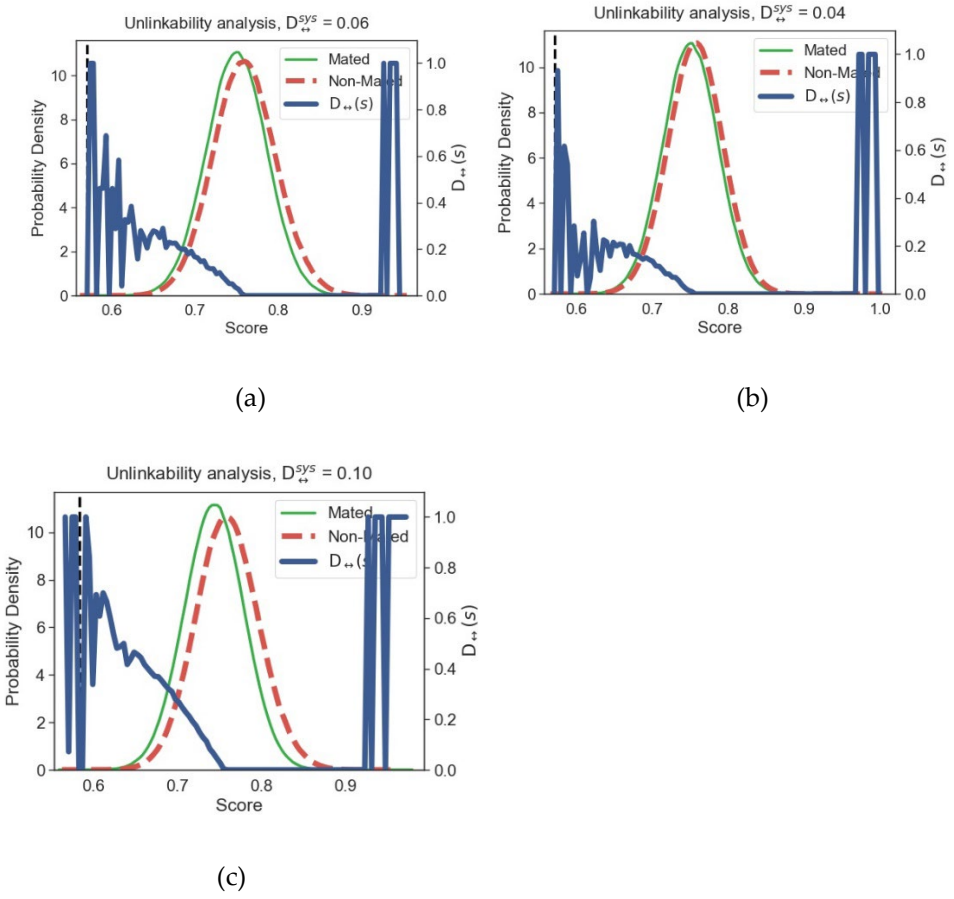


Figure 10. Unlinkability results for the proposed method on the (**a**) Pubfig, (**b**) Facescrub, and (**c**) YTF datasets.

5.2. Revocability Analysis

To analyze the revocability of the proposed scheme, we generated three score distributions: the mated-imposter score, the genuine score, and the imposter score [42]. The genuine and imposter score distributions were calculated by matching CSMoFN hashed vectors generated from the same and different subjects, respectively. The mated-imposter score is identical to the mated-samples score described in Section 5.1, and is calculated from the matching of two CB templates generated by the same subject with different user-specific seeds. In other words, it is assumed that the user revokes the old CSMoFN hashed vectors and creates a new instance, meaning that the mated-imposter score is the matching score of the old and new CSMoFN hashed vectors. The revocability criterion is deemed to be satisfied if the mated-samples score distribution overlaps with the imposter score distribution.

It can be observed from Figure 11 that the distributions of the mated-imposter and imposter scores substantially overlap for the three datasets, which indicates that the revocability criterion is satisfied.

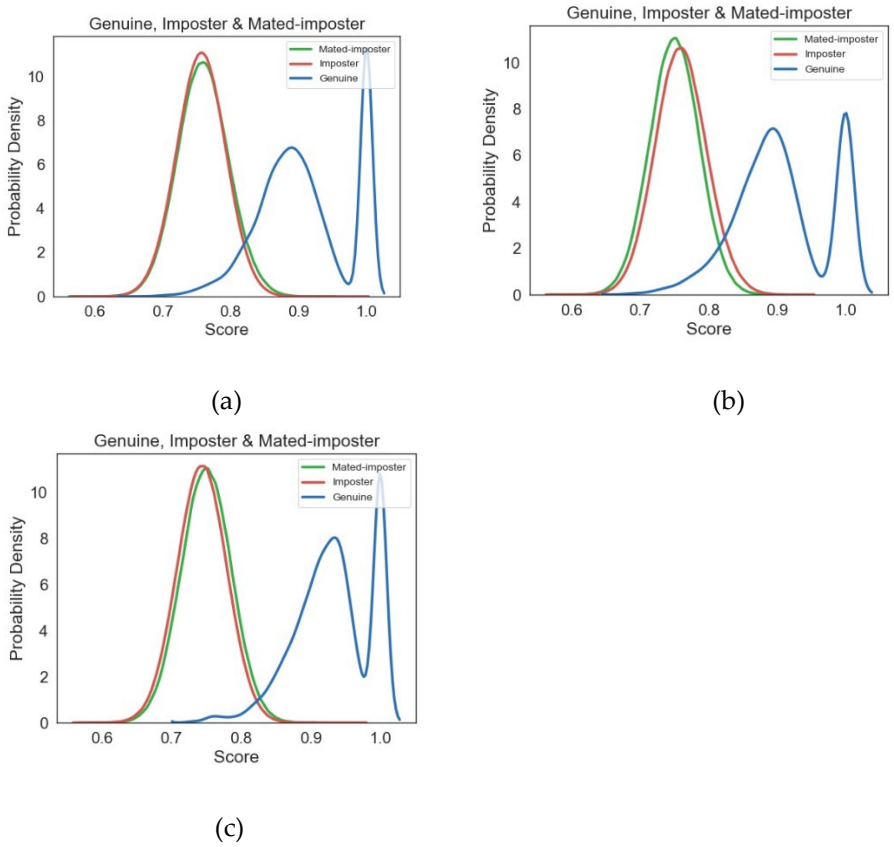


Figure 11. Revocability results from the proposed method on the (**a**) Ethnic, (**b**) IMDB Wiki, and (**c**) AR datasets.

6. Non-Invertibility Analysis

For our non-invertibility analysis, we consider two types of attack: brute-force and false acceptance (FA) attacks.

6.1. Brute-Force Attack

The goal of a brute-force attack is to estimate the CSMoFN hashed vectors by brute force, with it assumed that the attacker knows the structure of the CSMoFN and the corresponding hyperparameters [42].

The CSMoFN hashed vector s is a discrete vector with size k, where every element is within the range $[1, q]$. For a configuration such as $q = 32$ and $k = 512$ ($m = 4096$), the guessing complexity for each element is $q = 32 = 2^5$. Since there are k entries, the minimum guessing complexity is $2^{(5\times512)} = 2^{2560}$, which is prohibitively large in practice and prevents the attacker from going through all possible combinations. Furthermore, since CSMoFN is revocable, the hash vector can be replaced with a new one if it is found to be compromised.

6.2. False Acceptance Attack

An FA attack, also called a dictionary attack, is an attempt to gain illegal access to a biometric system [43]. This attack is realistic for any biometric system that relies on a decision threshold value. In other words, if the matching score of the authentication instance s^a and the transformed template s is less than a pre-defined threshold value τ, the right to access the biometric system is obtained. In the stolen token scenario, s^a is a CSMoFN hashed vector generated with a biometric vector $\boldsymbol{X}^a$ and stolen user-specific keys. The decision rule for authentication is then $Dist_H(s^a, s) > \tau$, where $Dist_H()$ is the Hamming distance and τ is the threshold value when the False Acceptance Rate (FAR) is equal to the False Rejection Rate (FRR).

To mitigate the FA attack, the threshold value should be set high, to achieve FAR = 0%. However, this implies a GAR (Genuine Acceptance Rate) reduction that suggests a degradation in accuracy. To balance the performance with security, a suitable threshold value τ should be carefully calibrated.

In this paper, the distance between the fake template s^* and s is calculated via $Dist_H(s, s^*) = UB_{imp}$, where UB_{imp} is the upper bound on the imposter scores for the considered dataset, and represents the worst scenario. To succeed in this approach, an attacker can attempt to find s^* to satisfy $Dist_H(s^a, s^*) = \tau$. That is, the goal is to generate a fake template such that the distance score with s^a falls into the interval $[\tau, UB_{imp}]$. Hence, the complexity of an FA attack can be estimated as $q^{m(\tau - UB_{imp})}$.

To analyze and respond to FA attacks in our context, it is necessary to determine the threshold value according to each GAR. Table 8 shows the complexity of an FA attack calculated for the six datasets used with the proposed method, with m = 4096 and q = 64. The complexity of the FA attack is evaluated based on GAR = 85%, 90%, and 95%. The UB_{imp} for each dataset can be obtained from the largest value of the imposter scores. Note that if $(\tau - UB_{imp})$ is negative, the complexity of the attack cannot be estimated.

In summary, the complexity of an FA attack can be increased in two ways: by increasing the value of m or reducing the GAR. However, the latter also implies a compromise in the accuracy performance. For GAR = 90%, our proposed method is reasonably robust in resisting FA attacks. In addition, an FA attack can be prevented by restricting the number of attempts at the authentication stage.

Table 8. Complexity of the false acceptance attack on the proposed system.

Datasets	q	m	UB_{imp}	τ	q^m	$(\tau - UB_{imp})$	Total Attack Complexity	GAR
AR	2^6	512	0.783	0.821	$2^{6\times512} = 2^{3072}$	0.038	$\approx 2^{117}$	**85%**
Ethnic			0.771	0.805		0.034	$\approx 2^{104}$	
Facescrub			0.789	0.824		0.035	$\approx 2^{108}$	
IMDB Wiki			0.792	0.817		0.025	$\approx 2^{77}$	
Pubfig			0.784	0.811		0.027	$\approx 2^{83}$	
YTF			0.785	0.818		0.033	$\approx 2^{101}$	
AR	2^6	512	0.783	0.807	$2^{6\times512} = 2^{3072}$	0.024	$\approx 2^{74}$	**90%**
Ethnic			0.771	0.790		0.019	$\approx 2^{58}$	
Facescrub			0.789	0.811		0.022	$\approx 2^{68}$	
IMDB Wiki			0.792	0.805		0.013	$\approx 2^{40}$	
Pubfig			0.784	0.798		0.014	$\approx 2^{43}$	
YTF			0.785	0.802		0.017	$\approx 2^{52}$	
AR	2^6	512	0.783	0.794	$2^{6\times512} = 2^{3072}$	0.011	$\approx 2^{34}$	**95%**
Ethnic			0.771	0.785		0.014	$\approx 2^{43}$	
Facescrub			0.789	0.803		0.014	$\approx 2^{48}$	
IMDB Wiki			0.792	0.787		-0.005	N/A	
Pubfig			0.784	0.786		0.002	$\approx 2^{6}$	
YTF			0.785	0.791		0.006	$\approx 2^{18}$	

7. Conclusions

In this paper, we propose a deep-learning-based multimodal cancelable biometrics scheme which we call CSMoFN. Our scheme fuses two biometric traits, namely the face and the periocular region, and is composed of three modules: a feature extraction and fusion module, a PSMoT module, and an MDC module. CSMoFN is trained by minimizing the ArcFace loss and the pairwise angular loss. Experiments were conducted on six datasets, with the verification performance approximately preserved with respect to its original counterpart. In addition, we have analyzed four conditions for our cancelable biometrics scheme and have shown that the proposed method satisfies them. In future research, we will consider more than two biometric modalities.

Author Contributions: Conceptualization, J.K. and A.B.J.T.; methodology, J.K.; software, J.K. and Y.G.J.; validation, J.K. and A.B.J.T.; formal analysis, J.K.; investigation, J.K.; resources, J.K. and Y.G.J.; data curation, J.K.; writing—original draft preparation, J.K.; writing—review and editing, J.K. and A.B.J.T.; visualization, J.K.; supervision, A.B.J.T.; project administration, A.B.J.T.; funding acquisition, A.B.J.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from the National Research Foundation of Korea (NRF), funded by the Korean government (MSIP) (no. NRF-2019R1A2C1003306).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets (AR dataset) were analyzed in this study. This data can be found here: https://www2.ece.ohio-state.edu/~aleix/ARdatabase.html (accessed on 10 February 2022). The data presented in this study are openly available in Ethnic at https://doi.org/10.1109/ICB45273.2019.8987278, reference number [36].The data presented in this study are openly available in Facescrub at https://doi.org/10.1109/ICIP.2014.7025068, reference number [37].The data presented in this study are openly available in IMDB Wiki at https://doi.org/10.1109/iccvw.2015.41, reference number [38].The data presented in this study are openly available in Pubfig at https://doi.org/10.1109/iccv.2009.5459250, reference number [39].The data presented in this study are openly available in YTF at https://doi.org/10.1109/CVPR.2011.5995566, reference number [40].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jain, A.K.; Nandakumar, K.; Nagar, A. Biometric Template Security. *EURASIP J. Adv. Signal Process.* **2008**, *2008*, 1–17. [CrossRef]
2. Jain, A.K.; Ross, A.; Prabhakar, S. An Introduction to Biometric Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2004**, *14*, 4–20. [CrossRef]
3. Ratha, N.K.; Chikkerur, S.; Connell, J.H.; Bolle, R.M. Generating Cancelable Fingerprint Templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 561–572. [CrossRef] [PubMed]
4. Jain, A.K.; Ross, A.A.; Nandakumar, K. *Introduction to Biometrics*; Springer Science & Business Media: New York, NY, USA, 2011.
5. Lahat, D.; Adali, T.; Jutten, C. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. *Proc. IEEE* **2015**, *103*, 1449–1477. [CrossRef]
6. Oloyede, M.O.; Hancke, G.P. Unimodal and Multimodal Biometric Sensing Systems: A Review. *IEEE Access* **2016**, *4*, 7532–7555. [CrossRef]
7. Canuto, A.M.P.; Pintro, F.; Xavier-Junior, J.C. Investigating Fusion Approaches in Multi-Biometric Cancellable Recognition. *Expert Syst. Appl.* **2013**, *40*, 1971–1980. [CrossRef]
8. Pinto, J.R.; Cardoso, J.S.; Correia, M.V. Secure Triplet Loss for End-to-End Deep Biometrics. In Proceedings of the 2020 8th International Workshop on Biometrics and Forensics (IWBF), Porto, Portugal, 29–30 April 2020; pp. 1–6.
9. Ding, C.; Tao, D. Robust Face Recognition Via Multimodal Deep Face Representation. *IEEE Trans. Multimed.* **2015**, *17*, 2049–2058. [CrossRef]
10. Al-Waisy, A.S.; Qahwaji, R.; Ipson, S.; Al-Fahdawi, S.; Nagem, T.A. A multi-biometric iris recognition system based on a deep learning approach. *Pattern Anal. Appl.* **2018**, *21*, 783–802. [CrossRef]
11. Alay, N.; Al-Baity, H.H. Deep Learning Approach for Multimodal Biometric Recognition System Based on Fusion of Iris, Face, and Finger Vein Traits. *Sensors* **2020**, *20*, 5523. [CrossRef]
12. Gunasekaran, K.; Raja, J.; Pitchai, R. Deep Multimodal Biometric Recognition Using Contourlet Derivative Weighted Rank Fusion with Human Face, Fingerprint and Iris Images. *Autom. J. Control. Meas. Electron. Comput. Commun.* **2019**, *60*, 253–265. [CrossRef]
13. Tiong, L.C.O.; Kim, S.T.; Ro, Y.M. Implementation of Multimodal Biometric Recognition Via Multi-Feature Deep Learning Networks and Feature Fusion. *Multimed. Tools Appl.* **2019**, *78*, 22743–22772. [CrossRef]
14. Algashaam, F.; Nguyen, K.; Banks, J.; Chandran, V.; Do, T.-A.; Alkanhal, M. Hierarchical Fusion Network for Periocular and Iris by Neural Network Approximation and Sparse Autoencoder. *Mach. Vis. Appl.* **2020**, *32*, 15. [CrossRef]
15. Luo, Z.; Li, J.; Zhu, Y. A Deep Feature Fusion Network Based on Multiple Attention Mechanisms for Joint Iris-Periocular Biometric Recognition. *IEEE Signal Process. Lett.* **2021**, *28*, 1060–1064. [CrossRef]
16. Jung, Y.G.; Low, C.Y.; Park, J.; Teoh, A.B.J. Periocular Recognition in the Wild With Generalized Label Smoothing Regularization. *IEEE Signal Process. Lett.* **2020**, *27*, 1455–1459. [CrossRef]
17. Soleymani, S.; Torfi, A.; Dawson, J.; Nasrabadi, N.M. Generalized Bilinear Deep Convolutional Neural Networks for Multimodal Biometric Identification. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 763–767.
18. Gomez-Barrero, M.; Rathgeb, C.; Li, G.; Ramachandra, R.; Galbally, J.; Busch, C. Multi-Biometric Template Protection Based on Bloom Filters. *Inf. Fusion* **2018**, *42*, 37–50. [CrossRef]

19. Jeng, R.-H.; Chen, W.-S. Two Feature-Level Fusion Methods with Feature Scaling and Hashing For Multimodal Biometrics. *IETE Tech. Rev.* **2017**, *34*, 91–101. [CrossRef]
20. Yang, W.; Wang, S.; Hu, J.; Zheng, G.; Valli, C. A Fingerprint and Finger-Vein Based Cancelable Multi-Biometric System. *Pattern Recognit.* **2018**, *78*, 242–251. [CrossRef]
21. Lee, M.J.; Teoh, A.B.J.; Uhl, A.; Liang, S.N.; Jin, Z. A Tokenless Cancellable Scheme for Multimodal Biometric Systems. *Comput. Secur.* **2021**, *108*, 102350. [CrossRef]
22. Gupta, K.; Walia, G.S.; Sharma, K. Novel Approach for Multimodal Feature Fusion to Generate Cancelable Biometric. *Vis. Comput.* **2021**, *37*, 1401–1413. [CrossRef]
23. Abdellatef, E.; Ismail, N.A.; Abd Elrahman, S.E.S.E.; Ismail, K.N.; Rihan, M.; Abd El-Samie, F.E. Cancelable Multi-Biometric Recognition System Based on Deep Learning. *Vis. Comput.* **2020**, *36*, 1097–1109. [CrossRef]
24. Talreja, V.; Valenti, M.C.; Nasrabadi, N.M. Deep Hashing for Secure Multimodal Biometrics. *IEEE Trans. Inf. Forensics Secur.* **2020**, *16*, 1306–1321. [CrossRef]
25. Sudhakar, T.; Gavrilova, M. Deep Learning for Multi-Instance Biometric Privacy. *ACM Trans. Manag. Inf. Syst. (TMIS)* **2020**, *12*, 1–23. [CrossRef]
26. El-Rahiem, B.A.; Amin, M.; Sedik, A.; Samie, F.E.; Iliyasu, A.M. An efficient multi-biometric cancellable biometric scheme based on deep fusion and deep dream. *J. Ambient. Intell. Humaniz. Comput.* 2021, *in press*. [CrossRef] [PubMed]
27. Teoh, A.B.J.; Cho, S.; Kim, J. Random Permutation Maxout Transform for Cancellable Facial Template Protection. *Multimed. Tools Appl.* **2018**, *77*, 27733–27759. [CrossRef]
28. Wang, J.; Zhang, T.; Song, J.; Sebe, N.; Shen, H.T. A Survey on Learning to Hash. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 769–790. [CrossRef]
29. Du, H.; Shi, H.; Zeng, D.; Mei, T. The Elements of End-to-End Deep Face Recognition: A Survey of Recent Advances. *arXiv* **2009**, arXiv:2009.13290. [CrossRef]
30. Guo, Y.; Zhang, L.; Hu, Y.; He, X.; Gao, J. Ms-celeb-1m: A Dataset and Benchmark for Large-Scale Face Recognition. In Proceedings of the European Conference on Computer Vision, Amstedrdam, The Netherlands, 11–14 October 2016; Leibe, B., Matas, J., Sebe, N., Eds.; Max Welling Springer: Cham, Switzerland; pp. 87–102.
31. Lee, H.; Low, C.Y.; Teoh, A.B.J. SoftmaxOut Transformation-Permutation Network for Facial Template Protection. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021.
32. Li, W.; Zhang, S. Binary Random Projections with Controllable Sparsity Patterns. *arXiv* **2020**, arXiv:2006.16180 [cs, stat].
33. Jin, A.T.B. Cancellable biometrics and multispace random projections. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), New York, NY, USA, 17–22 June 2006; p. 164.
34. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. Arcface: Additive Angular Margin Loss for Deep Face Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 4690–4699.
35. Martinez, A.; Benavente, R. *The AR Face Database: CVC Technical Report, 24*; Centre de Visioper Computador Universitat Aut onoma de Barcelona: Barcelona, Spain, 1998.
36. Tiong, L.C.O.; Teoh, A.B.J.; Lee, Y. Periocular Recognition in the Wild with Orthogonal Combination of Local Binary Coded Pattern in Dual-Stream Convolutional Neural Network. In Proceedings of the 2019 International Conference on Biometrics (ICB), Crete, Greece, 4–7 June 2019.
37. Ng, H.-W.; Winkler, S. A Data-Driven Approach to Cleaning Large Face Datasets. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014.
38. Rothe, R.; Timofte, R.; Van Gool, L. Dex: Deep Expectation of Apparent Age from a Single Image. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 7–13 December 2015; pp. 10–15.
39. Kumar, N.; Berg, A.C.; Belhumeur, P.N.; Nayar, S.K. Attribute and simile classifiers for face verification. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 365–372.
40. Wolf, L.; Hassner, T.; Maoz, I. Face Recognition in Unconstrained Videos with Matched Background Similarity. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 529–534.
41. Gomez-Barrero, M.; Galbally, J.; Rathgeb, C.; Busch, C. General Framework to Evaluate Unlinkability in Biometric Template Protection Systems. *IEEE Trans. Inf. Forensics Secur.* **2017**, *13*, 1406–1420. [CrossRef]
42. Jin, Z.; Hwang, J.Y.; Lai, Y.-L.; Kim, S.; Teoh, A.B.J. Ranking-Based Locality Sensitive Hashing-Enabled Cancelable Biometrics: Index-of-Max Hashing. *IEEE Trans. Inf. Forensics Secur.* **2017**, *13*, 393–407. [CrossRef]
43. Tams, B.; Mihăilescu, P.; Munk, A. Security Considerations in Minutiae-Based Fuzzy Vaults. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 985–998. [CrossRef]