THIS IS A PRE-PRINT VERION OF THE CURRENT STUDY: Is Warping-based Cancellable Biometrics (still) Sensible for Face Recognition ?

Simon Kirchgasser[§], Andreas Uhl
University of Salzburg
Jakob-Haringer-Str. 2, Salzburg, AUSTRIA
{skirch, uhl}@cs.sbg.ac.at

Yoanna Martínez-Díaz[§], Heydi Mendez-Vazquez
Advanced Technologies Application Center
(CENATAV), Havana, CUBA
{ymartinez, hmendez}@cenatav.co.cu

Abstract

We conduct an ISO/IEC Standards 24745 and 30136 compliant assessment of block-based warping sample transformation techniques aiming for template protection. Particular focus is laid on the results' evaluation considering the evolution of face recognition technology ranging from more "historic" hand-crafted features to state-of-the-art deep-learning (DL) based schemes. It turns out that the high robustness of todays face recognition technology can handle the geometrical distortions as introduced by warping as another form of variability like pose, illumination, and expression variations, thereby disabling the intended protection functionality of warping. Therefore, block-based warping sample transformation must not be used as template protection technique for todays state-of-the-art face recognition schemes, while some settings could be identified providing template protection to some extent for less recent face recognition technology.

1. Introduction

Face recognition is applied in many different scenarios like forensics, surveillance or border control. In most cases the used systems have to deal with natural but uncontrolled environmental conditions including variations in pose, illumination and facial expressions [6, 10, 18]. Recently, face recognition techniques have evolved with the use of DL techniques being able to deal with the appearance of these natural variations, significantly boosting recognition accuracy. However, it is not only the robustness against various variations that is important. The ability to protect the acquired subject's biometric information and thereby preserving privacy is increasingly attracting attention. There are various possibilities in face biometrics how the subject's privacy can be protected, e.g. [12, 13]. All these – so called template protection schemes – can be categorized into several classes [27]. One important class is based on non-invertible transformations and is termed *cancelable biometrics*. This class of template protection schemes has been introduced in a seminal paper by Ratha et al. [29] focusing on block re-mapping for fingerprints (in the minutiae domain) and warping for facial data (in the image domain). These techniques have also been proposed and thoroughly evaluated later on iris biometrics [17, 23], finger vein [21, 28] databases, and face images in video surveillance data [22], respectively.
In the current study, we re-visit the application of image domain block-based warping [31] to protect facial images in the context of face recognition for the following reasons:

(i) Since the introduction of the approach in [29] 18 years back, no systematic assessment according to ISO/IEC Standards 24745 and 30136 [1, 2] requirements has been conducted, at least not in the context of face recognition. In the context of iris recognition an related security analysis was presented on Bloom Filters [5], while security aspects of cancelable iriscodes based on a secret permutation were discussed in [4].

(ii) Deep-learning based face recognition systems have significantly improved robustness against various variations in acquisition conditions as mentioned above. Therefore, we suspect that geometrical variations as introduced by block-based warping might be treated by this more recent class of recognition schemes as just another acquisition condition variation, thereby disabling the protection aim due to increased robustness.

(iii) The application of warping-based non-invertible transformations is computationally efficient (which is of course desirable to keep the computational costs low).

(iv) The application in the image domain (as opposed to using template protection techniques in the feature / template domain) is an elegant and highly flexible solution to the template protection problem. On the one

[§] These authors contributed equally to this work as first authors.

hand, the ”original (real) template” is never generated and thus cannot be compromised in any stage of the recognition scheme’s operation. On the other hand, this approach can be applied to almost any recognition scheme including template-less systems, where enrollment and authentication sample data similarity is learned in pairs (or triplets) without explicitly storing templates.

To address these issues, we conduct an experimental study focusing on ISO/IEC Standard 24745/30136 compliant security evaluation, in particular relating results of several recent deep neural network-based face recognition systems [7, 8, 24] to more traditional and well-established ones.
The rest of this paper is organized as follows: In Section 2 we present a compact literature review on the applied template protection technique including several properties which need to be fulfilled according to ISO/IEC Standard 24745/30136 and are evaluated in the experimental Section 4. Subsequently, the used face recognition schemes, the dataset and the experimental set-up are explained (see Section 3). The experimental evaluation regarding recognition performance and unlinkability aspects is presented in Section 4. Finally, Section 5 concludes this paper along with an outlook on future work.

2. Related Work

A well-defined biometric template protection scheme needs to fulfil four requirements which are defined in ISO/IEC Standard 24745 [1]: *Non-invertibility or Irreversibility*, *Revocability or Renewability*, *Non-linkability or Unlinkability* and *Performance preservation*. Revocability ensures that a compromised template can be revoked without exposing biometric information, i.e. the original biometric trait (template) remains unaltered and is not compromised. In case of the applied warping technique this property is given by the design of the method. A new protected template can be generated by selecting a new key which defines the block size of the regular input grid and further controls the amount of distortion introduced to the unprotected biometric information. Thus, after removing the compromised data, a new protected template representing the same biometric instance can be generated easily. The aspect of performance preservation shall ensure that applying a certain protection scheme does not lead to a significant recognition performance degradation of the used recognition system. This latter aspect is experimentally evaluated and the corresponding results will be presented in Section 4.

Irreversibility Irreversibility as defined in ISO/IEC Standard 24745 [1] shall ensure that it is not possible to extract the original biometric information from the protected biometric template, even in case the key to produce the protected template is available. In [16] the authors describe brute force attacks, reconstruction attacks and Hamming weights attacks as potential threats that can break the irreversibility of a protected biometric template. Furthermore, they postulate that it is necessary to be aware that a successful attack breaking the irreversibility also breaks the unlinkability. The application of warping to protect a biometric face trait fulfils the irreversibility property by design depending on the selected parameters. If the chosen warping parameters introduce a high amount of distortions, as visible in Figure 1 e), the resulting interpolation effects are assumed to result in high irreversibility. However, we do not conduct explicit experiments on this issue as explained when introducing the evaluation protocol (Section 3.3) and discussing the results of the other experiments (Section 4).

Unlinkability The property of unlinkability is meant to guarantee that protected biometric templates can not be linked across various applications or databases. The ISO/IEC Standard 24745 [1] defines templates to be *fully linkable* if a method exists which is able to decide if two templates protected using a different key were extracted from the same biometric sample with a certainty of 100%. The degree of linkability depends on the certainty of the method which decides if two protected templates originate from the same capture subject. Unfortunately, the standard only defines what the unlinkability property shall ensure but gives no generic way of quantifying it. Gomez et al. [15] introduced a framework to evaluate the unlinkability of a biometric template protection system based on the comparison scores which we adopt also for this work. The authors proposed a global measure as evaluation method - the so called D_{sys} measurement. This D_{sys} measure originally ranges from 0 to 1. We have transformed the range $[0, 1]$ to $[0, 100]$ for readability reasons. Thus, 0 represents the best achievable unlinkability score (fully unlinkability), while a D_{sys} value of 100 indicates full linkability. Furthermore, it was stipulated in the same work [15] that at least 10 different keys should be considered during the unlinkability analysis. Thus, we have also selected 10 different keys for our performance and unlinkability analysis (Section 4).

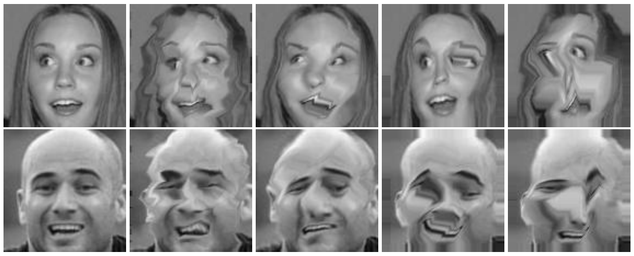
Key-Selection: System vs. Subject-Specific Finally, it is necessary to discuss another important aspect which is not covered in the ISO/IEC Standard 24745 [1]: What is better? Selecting a subject-specific or a system key? In the subject-specific key approach, the template of each subject is generated by a key which is distinct for each subject while for a system key, the templates of all subjects are generated by the same key.
Subject-specific keys have advantages in terms of preserving the subjects’ privacy compared to system keys. Assigning an individual key to each subject ensures that if

an adversary gets to know the key of one of the subjects, the entire database can not be compromised as each key is individual. A subject key further ensures that insider attacks performed by legitimate registered subjects can not be done that easily. This potential attack involves a registered subject, who has access to the system and to the template database. This adversary subject wants to be legitimated as one of the other subjects of the same biometric system by copying one of the own templates over templates belonging to another capture subject. Thus, an adversary can claim this identity as well and might be authenticated as a genuine subject. In case subject keys are used, such a copying process is not straight forward and costly as each of the templates stored in the database has been generated using an individual key. Thus, it might be easier for an advisory to create a new genuine subject exhibiting the wanted biometric information protected by a new key which is set in order to get the legitimation as wanted. Another advantage of subject keys is that it is likely that the system's recognition performance is enhanced as more inter-subject variability is introduced. The additional variability in combination with the variations between different biometric subjects could enable a better separation of genuines and impostors enhancing the overall system's performance. Advantages of the system-specific key approach include the easier revocation of compromised templates as the generation of a new version of all templates only includes a single key and the much simpler key-management requirements.

3. Methodology

3.1. Block-Based Warping

"Block-based Warping" (originally termed *mesh warping* [31] or shortly "warping") applies a function to each pixel in the given image which maps the pixel of the input at a given position to a certain position in the output. Thus, a pixel can also remain at its original position. The applied mapping defines a new image or template (depending on the domain where the mapping function is applied to) containing the same information as the original input but in a distorted representation. The warping approach employed in this study is a combination of using a regular grid (defined by non-overlapping blocks) and a distortion function based on spline interpolation which results in a geometrical deformation (variation) of the input image. The regular grid is deformed per each block and adjusted to the warped output grid. The number of blocks in the output is the same as in the input, but the content of each individual block is distorted in the warped output.
The distortion is introduced by randomly altering the edge positions of the regular grid, leading to a non-predictable deformation of this grid, the warped output grid. Spline based interpolation of the input information is applied to adopt the area of each block with respect to the smaller or larger block area obtained after the deformation application. Thus, warping might either stretch or shrink the area of the block depending on the varied edge positions. The introduced distortion is key dependent and the key defines the seed value for the random generator responsible for the replacement of the grid edges (maximum pixel offset parameter). The key further contains the size of each block given in the regular grid. Examples face images depicting different amounts of distortions (ensuring the described properties) are shown in Figure 1. The first number in the naming of the given example image (b) - (e) presents the used block size while the second number defines the maximal used pixel offset (further information on the chosen parameters is given in Section 3.3). For more details about different warping approaches the interested reader is referred to [14].



(a) original (b) warp_8_4 (c) warp_16_6 (d) warp_20_9 (e) warp_20_25

Figure 1: Example images displaying various distortions introduced to an original image by using the warping scheme.

3.2. Face Recognition and Dataset Description

Several techniques exist in face biometrics to perform verification or identification tasks which have been evaluated thoroughly in various studies, e.g. [20]. Thus, we only briefly describe methods including well-established traditional approaches and more recent ones which we selected as face recognition schemes. First, we consider methods based on traditional handcrafted local descriptors such as Local Binary Patterns (LBP) [3] and Multi-Block LBP (MBLBP) [32]. Both selected descriptors were extracted by using cells regions of size 14×14. The final feature vector, representing the face images, is a histogram containing all single histograms of each cell region extracted before. Secondly, we evaluated two learning-based local descriptors based on the Fisher Vector representation. Specifically, we tested the Video Fisher Vector Faces (VF^2) descriptor [26] that encodes SIFT features and the Logistic Binary Video Fisher Vector Faces (LBinVF^2) [25], which efficiently encodes so called BRIEF descriptors. Finally, we applied three recent deep convolutional neural networks including ResNet-ArcFace (ArcFace) [8], MobileFaceNet (MobileFace) [7] and ShuffleFaceNet (ShuffleFace) [24] as all three methods resulted in good performance values, which have been presented in previous studies. In order to compute the

comparison scores (done by the usage of a support vector machine), once the features are extracted, we use the cosine distance as the similarity measure for the DL networks and the Fisher Vector approaches. In case of the handcrafted descriptor based methods a chi-squared distance measure was selected.
The Labeled Faces in the Wild (LFW) database [19] is well known as a public benchmark for unconstrained face verification. It contains $13,233$ web-collected face images from $5,749$ different identities, with large variations in pose, expression and illuminations. $6,000$ face pairs are divided into 10 non-repeating subsets of images pairs. Thus, each subset includes 300 positive pairs (images from the same person) and 300 negative pairs (images from different people). In the following we will name the positive pairs 'genuine' and the negative pairs 'impostor'. All face images were aligned and cropped to the size of 112×112 by using the RetinaFace detector [9] before subsequently applying template protection (warping) followed by one of the described recognition schemes.

3.3. Evaluation protocol

In the scope of this study several experiments (A, B, C, D) have been conducted. For all these experiments the necessary comparison scores have been obtained by using the same protocol. The applied protocol was suggested to be used for evaluations using the LFW database and has been used in previous work, e.g. [19]. According to the dataset structure (10 subsets containing 300 genuine pairs and 300 impostor pairs each) and the corresponding protocol a total of 3000 genuine and 3000 impostor scores for each experiment can be obtained. The performance is reported as 10-fold cross validation by the mean accuracy (*acc* in percent) of the Support Vector Machine classifier as it was suggested in [19].

(A) **Baseline:** At first, all selected face recognition methods have been applied and assessed on the original face images to achieve baseline recognition performance results. These baseline results are presented in the second column of Table 1 and Table 1 to allow a easier comparison to the additional experiments (B) and (C).

After the baseline was established, the described warping technique was applied to each face image using four different parameter settings. Each setting consists of block size and offset parameter, controlling the amount of distortion introduced to the image. We have selected a block size of 8×8, 16×16 and 20×20 pixels, while maximal offset values of 4, 6, 9 and 25 have been chosen. The settings are abbreviated by warp_8_4, warp_16_6, warp_20_9 and warp_20_25, where the first value describes the block size and the second number the offset. The first setting using an offset of 4 pixel is describing slight perturbation only, while an offset of 6 and 9 leads to an average facial modification for the LFW database. The final offset choice of 25 pixel results in an extreme and strong face image distortion. Furthermore, we have selected both subject-specific and system keys. We have conducted two different types of experiments (B and C) using the warped face images:

(B) **Performance Preservation:** The second experiment compares warped facial images against warped facial images with identical warping parameters. The results should highlight the capability of how well the applied face recognition systems can perform comparisons in the warped image domain, thus performance preservation is assessed in this manner. The averaged results of this experimental evaluation considering 10 repetitions to allow unlinkability experiments (D) are presented in Table 1.

(C) **Protection Strength:** The third experiment is defining a kind of special case of unlinkability and irreversibility evaluation. On the one hand, if an attacker has access to a biometric database of potential candidates, stored as unprotected samples, this allows the opportunity to decide which sample corresponds to a given protected sample, and thus it can be interpreted as a partial irreversibility attack. On the other hand, if the accuracy reported by the protection strength experiments is high than a comparison algorithm can be utilized as a method to distinguish mated and non-mated samples as it is intended by the following unlinkability experiments.
To perform this protection strength analysis ee compare original undistorted samples against warped ones to assess protection strength (using subject and system-specific keys). This type of conducted experiments aims at showing how good the privacy of a subject is protected. The higher the resulting accuracy is, the worse is the protection. In case sensible recognition results can be achieved in this setting, protection strength is obviously low as subjects can be recognized although we are comparing original to protected data. Simultaneously, the robustness of different face recognition types against the artificial variations introduced is evaluated. Similar to the performance preservation experiments (B) the averaged acc. results (considering once more 10 repetitions) including the standard deviation are shown in Table 2.

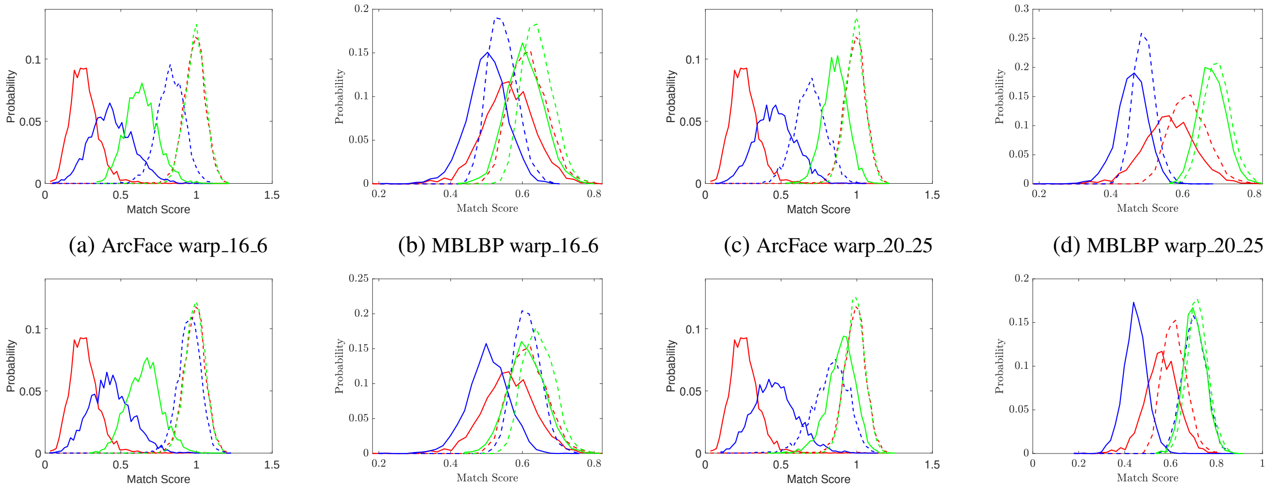
(D) **Unlinkability:** Finally, we also investigated the aspect of unlinkability by comparing images which have been warped by the use of different keys. Thus, to be inline with the protocol introduced by [15] it was necessary to repeat the warping 10 times for each parameter

method	baseline	warp_8_4		warp_16_6		warp_20_9		warp_20_25	
		subject	*system*	*subject*	*system*	*subject*	*system*	*subject*	*system*
ArcFace	**99.87 ± 0.2**	**98.20 ± 0.2**	**94.35 ± 1.0**	**98.80 ± 0.1**	**94.87 ± 1.8**	**98.33 ± 0.2**	**93.00 ± 3.0**	91.79 ± 0.7	**79.93 ± 7.5**
MobileFace	99.72 ± 0.3	98.09 ± 0.1	91.34 ± 1.3	98.49 ± 0.1	91.30 ± 2.7	97.91 ± 0.2	88.95 ± 4.0	93.85 ± 0.5	77.81 ± 5.5
ShuffleFace	99.62 ± 0.4	97.79 ± 0.2	90.98 ± 1.4	98.18 ± 0.2	90.17 ± 2.7	97.32 ± 0.2	87.87 ± 4.3	91.42 ± 0.7	76.71 ± 5.5
VF^2	76.57 ± 1.3	87.71 ± 1.2	74.10 ± 0.3	90.70 ± 0.3	74.26 ± 0.6	92.91 ± 0.3	73.49 ± 0.9	94.15 ± 2.7	73.10 ± 1.0
LBinVF^2	73.62 ± 1.0	76.34 ± 0.5	72.41 ± 0.4	79.84 ± 0.4	72.60 ± 0.6	83.09 ± 0.2	72.05 ± 0.9	84.98 ± 3.6	71.56 ± 0.9
MBLBP	67.02 ± 1.9	83.73 ± 0.2	67.55 ± 0.3	89.01 ± 0.2	67.29 ± 0.5	93.32 ± 0.2	67.69 ± 0.8	**99.52 ± 0.2**	66.79 ± 1.0
LBP	66.41 ± 1.7	83.60 ± 0.3	66.95 ± 0.6	88.75 ± 0.4	66.71 ± 0.5	93.45 ± 0.2	66.60 ± 0.6	99.46 ± 0.1	65.91 ± 1.1

Table 1: Recognition Preservation results (B): Face verification results (mean acc. and standard deviation in percent) for the applied warping template protection scheme using subject-specific and system keys.

method	baseline	warp_8_4		warp_16_6		warp_20_9		warp_20_25	
		subject	*system*	*subject*	*system*	*subject*	*system*	*subject*	*system*
ArcFace	**99.87 ± 0.2**	**96.95 ± 0.2**	**96.83 ± 0.4**	**96.58 ± 0.2**	**96.77 ± 1.8**	**93.87 ± 0.3**	**95.34 ± 2.4**	**72.95 ± 1.3**	**78.37 ± 8.8**
MobileFace	99.72 ± 0.3	96.15 ± 0.3	96.02 ± 0.7	95.42 ± 0.2	95.30 ± 2.3	91.95 ± 0.3	93.20 ± 3.1	72.78 ± 1.4	76.80 ± 8.6
ShuffleFace	99.62 ± 0.4	95.68 ± 0.2	95.61 ± 0.7	94.72 ± 0.2	94.64 ± 2.5	91.17 ± 0.3	92.16 ± 3.3	70.71 ± 1.1	74.70 ± 8.7
VF^2	76.57 ± 1.3	74.96 ± 0.9	75.41 ± 0.4	75.19 ± 0.3	75.37 ± 0.3	74.18 ± 0.4	74.44 ± 0.8	72.17 ± 3.1	73.15 ± 2.1
LBinVF^2	73.62 ± 1.0	72.18 ± 0.3	72.39 ± 0.3	71.87 ± 0.3	71.76 ± 0.6	70.30 ± 0.3	71.15 ± 1.0	68.47 ± 3.1	69.74 ± 2.1
MBLBP	67.02 ± 1.9	64.54 ± 0.4	64.58 ± 0.2	64.07 ± 0.5	64.19 ± 0.8	62.24 ± 0.5	63.19 ± 1.3	55.88 ± 0.7	57.53 ± 2.0
LBP	66.41 ± 1.7	63.21 ± 0.4	63.18 ± 0.4	63.29 ± 0.3	63.19 ± 0.7	60.95 ± 0.5	61.86 ± 1.3	54.47 ± 0.6	56.96 ± 1.5

Table 2: Protection Strength results (C): Face verification results (mean acc. and standard deviation in percent) by comparing original versus protected images using subject-specific and system keys.



(a) ArcFace warp_16_6

(b) MBLBP warp_16_6

(c) ArcFace warp_20_25

(d) MBLBP warp_20_25

(e) ArcFace warp_16_6

(f) MBLBP warp_16_6

(g) ArcFace warp_20_25

(h) MBLBP warp_20_25

Figure 2: Example images which display the genuine and impostor distributions behaviour using system keys (first row) and subject keys (second row).

setting using distinct keys. The unlinkability results shown in Table 3 are the result of an averaging process thus the averaged D_{sys} values are presented together with their standard deviation. Further details on the comparison protocol can be found in [15].

4. Experimental Results

For the baseline experiments (A) we found that DL-based face recognition scheme types work best on the considered LFW dataset. The applied DL methods (ArcFace, MobileFace and ShuffleFace) clearly outperformed all other schemes (acc. is almost 100%)) and are followed by techniques applying the Fisher Vector representation (acc. between 73% and 77%). Both LBP-based approaches performed worst (acc. around 67%). Overall, the ranking among the different recognition systems corresponds to our expectations and earlier results in literature impressively documenting the progress made over the last years.

Table 1 displays the results of our *performance preserva-*

method	warp_8_4		warp_16_6		warp_20_9		warp_20_25	
	subject	*system*	*subject*	*system*	*subject*	*system*	*subject*	*system*
ArcFace	80.55 ± 0.59	80.95 ± 1.75	75.57 ± 0.8	77.94 ± 4.8	63.82 ± 0.9	71.52 ± 5.7	17.13 ± 1.6	27.21 ± 10.3
MobileFace	72.52 ± 0.91	74.69 ± 2.35	61.16 ± 1.0	67.69 ± 5.9	48.04 ± 1.0	59.25 ± 6.1	18.99 ± 1.4	23.03 ± 8.1
ShuffleFace	70.79 ± 0.83	72.81 ± 2.24	58.63 ± 0.9	65.37 ± 5.4	46.80 ± 1.2	56.18 ± 6.1	15.95 ± 1.4	20.40 ± 7.7
VF^2	42.63 ± 0.46	43.72 ± 0.68	38.93 ± 0.6	41.77 ± 1.4	33.53 ± 0.8	39.35 ± 3.0	27.21 ± 6.1	34.07 ± 5.8
LBinVF^2	38.82 ± 0.44	39.02 ± 0.73	36.45 ± 0.5	36.92 ± 1.6	31.86 ± 0.7	35.05 ± 2.2	25.60 ± 6.1	29.79 ± 5.2
MBLBP	**28.59 ± 0.69**	**28.74 ± 0.70**	**25.06 ± 0.7**	**27.08 ± 1.5**	**17.29 ± 0.8**	**24.74 ± 2.7**	**6.62 ± 0.7**	**11.23 ± 3.2**
LBP	32.83 ± 0.61	32.53 ± 0.87	29.15 ± 0.8	31.06 ± 1.5	19.51 ± 0.9	28.04 ± 2.6	7.28 ± 1.0	13.83 ± 4.1

Table 3: Unlinkability results (D): Averaged D_{sys} unlinkability scores for the warped dataset using subject and system keys. The best results (low values, representing unlinkability) for each parameter setting are highlighted in bold numbers.

tion experiments (B). The results of applying recognition in the protected (i.e. warped) domain show a clear difference between the usage of subject-specific and system keys. For system keys, we observe a significantly decreasing recognition accuracy for increasing warping strength (thus, warp_20_25 resulted in the worst performance: Deep learning-based schemes loose about 15% accuracy and go down to almost 75%). However, this is only the case for recognition schemes with sensible accuracy in the baseline experiments, for weaker schemes this decrease is observed only to a lesser extent. As the comparison was done using different keys for probe and reference images, it is certain that the performance reduction is assumable based on a lower false-acceptance-rate (FAR) as the applied subject-key specific warping introduces a 2nd layer of separability as for each user the corresponding samples are secured with the same key while all other user's samples are protected by a different ones.

While DL based schemes are reduced in their accuracy down to 91% only (as compared to using system keys), for all non-DL methods even a clear accuracy enhancement is observed. In particular, the LBP schemes benefit a lot as their accuracy (baseline around 67%) increases to approximately 84% (warp_8_4) or even up to 99.52% (warp_20_25 using MBLBP). The application of warping using subject-specific keys enables a two factor authentication, which not only takes the biometric information present in the original unprotected images into account, but also the newly introduced geometrical variations which differ among users. Due to the higher robustness of the DL based schemes, this effect is not observed in their results (as the warping-related variations are compensated by the robustness properties).

For the DL based schemes, we assume that a re-training of the networks including warped images in the training dataset would result in a recognition accuracy increase (we only used networks trained on original unprotected face images). Overall, the property of performance preservation (B) as recommended by ISO/IEC Standard 24745 [1] is fulfilled only for subject-specific keys as we notice a partially significant accuracy drop for the system key approach. The results of subject-specific keys have to be taken with care as discussed in [30] for iris data. There it is stated that a biometric (iris) recognition system becomes highly vulnerable even in case the subject-specific key of only one subject is stolen, lost, shared or duplicated. This vulnerability might be similar for face recognition as well.

The results presented in Table 2 describe the recognition performance (accuracy) in case original unprotected images (e.g. in the gallery) are compared against protected ones (e.g. the probe images) - *protection strength* analysis (C). Contrasting to the evaluation of recognition performance preservation as done before, good accuracy is now a negative result – as we compare unprotected samples to protected ones, good recognition performance indicates that recognition is possible although probe samples are being protected.

Regardless if subject- or system-specific keys are taken into account it can be observed that a comparison between unprotected and protected images leads to an accuracy of clearly above 50% thus, recognition partially works despite the probe samples being protected. The worst accuracy is observed in case of warp_20_25 (the stronger the warping is applied, the more challenging the recognition task becomes). However, an accuracy of more than 70% in the warp_20_25 setting for the DL methods is a remarkable result if the strongly distorted images (see Figure 1) these results are based on are taken into account. This underpins the extremely high robustness of these techniques being partially able to deal with these strong geometrical distortions.

The accuracy of around 55% for the LBP-based techniques under warp_20_25 exhibits a result as required for successful template protection. Given the corresponding results for subject-specific keys with respect to recognition preservation, this parameter configuration is the only one which seems to be suited for actual template protection. A further aspect to be noted is that for average and strong warping parameters the system key approach results in better accuracy values, i.e. subject-specific keys introduce additional entropy enhancing protection. This observation is perfectly in-line with the results obtained for the performance preservation experiments (B). The usage of subject-specific keys introduces an additional (but only slight) protection layer to the protected images. The additional subject-specific geometric variations ensure that each subject is a little bit "more unique" than the original sample and system-key protected images and thus the applied face recognition systems tend to have more problems comparing original data with the protected ones. Summarizing, the outcomes of experiments (C) clearly indicate that an application of warping is not a sufficient privacy protection method, especially if DL recognition systems are used. Additionally, it must be noted that from Table 2 and the linking knowledge from unlinkability and irreversibility these experiments also indicate that the deployed template protection scheme will not achieve unlinkability.

The observations made so far are visually underpinned in Figure 2. In this figure the dashed lines correspond to the impostor score distributions, while the solid lines represent the genuine score distributions. The red color corresponds to the baseline experiments (A), while blue lines depict performance preservation (B) and green lines the protection strength (C) analysis, respectively.

The genuine and impostor score distributions for baseline experiments (A) show that only for the usage of DL face recognition schemes (see Figure 2 *(a)* and *(e)*) a clear separability of genuine and impostor scores can be described, regardless if system or subject keys are used.

What else do we expect in terms of relation between genuine and impostor distributions ? For decent performance preservation (B), blue genuine and impostor distributions should preferably exhibit no overlap but should at least show an overlap similar to the unprotected (red) case. In fact we observe for the DL-based cases (Figures 2 *(a)*, *(c)*, *(e)*, and *(g)*) the separation of blue distributions being worse as the red ones - thus, this corresponds to reduced recognition accuracy and confirms results of table 1. For the LBP-based schemes the overlap between genuine and impostor distribution is similar for experiments (A) and (B), except for the subject-specific keys, which clearly show a smaller overlap when comparing warped gallery and probe samples with identical parameters (also in perfect correspondence to table 1 results).

For a high protection strength (C), we expect green genuine distributions to have maximal overlap with the red and green impostor distributions (indicating that no genuine matches are recognized as such due to the protected samples). We are able to observe this almost perfect overlap for the LBP results (Figures 2 *(b)*, *(d)*, *(f)*, and *(h)*), the results for the DL-based techniques display a reduced but still large overlap for middle warping parameters (Figures 2 *(c)* and *(g)*)) but little overlap for the strong warping warp_20_25 (Figures 2 *(a)* and *(e)*). Thus, confirming the results of table 2, protection strength is very weak for DL based schemes, whereas it is acceptable for extreme warping parameters.

The outcomes for the unlinkability experiments (D) are presented in Table 3. According to the description given in Section 2, low D_sys values represent good unlinkability, while high values indicate a severe extent of linkability present among protected templates. Overall it can be stated that we only observe results with sensible unlinkability if warp_20_25 using MBLBP and LBP is considered. The absolute values suggest "semi-unlinkability" according to [15] and [11]. In particular, all DL-based methods are close to full linkability except for the warp_20_25 setting. Again, subject-specific keys provide better unlinkability as compared to the system key scenario.

This result underlines once more that warping-based sample transformation is not an appropriate template protection method especially if DL-based recognition schemes are applied. Only in case of applying very strong warping parameters and using MBLBP and LBP, the "oldest" recognition schemes investigated, the D_sys values indicate to fulfill the required property of unlinkability to some degree.

5. Conclusion

ISO/IEC Standard 24745 compliant assessment of block-based warping sample transformation techniques reveals that for state-of-the-art DL-based face recognition, protection strength and unlinkability are not sufficient and corresponding deployment is depreciated. Earlier face recognition schemes based on "handcrafted" LBP features turn out to benefit from this template protection approach when applying strong warping and subject-specific keys if focusing on the recognition accuracy only. In this setting the recognition accuracy is improved at the cost of the introduction of a second authentication factor (i.e. the key) and corresponding key management requirements. However, we also cannot really recommend warping as a protection measure for LBP as the unlinkability, although better than with DL-based schemes is already at risk.

For a complete assessment with respect to ISO/IEC Standard 24745, irreversibility needs to be considered. However, as the unlinkability and protection strength experiments already reveal that template comparison is already possible without actually reversing the warping transformation, there is no need to investigate irreversibility, as an approximation based on an inverted warping would bring the protected templates even closer to the originals. Furthermore, it was discussed in the setup of the protection strength experiments that a close relation to an irreversibility attack is given by performing these experiments. Thus, the combined view at unlinkability and protection strength results implies that irreversibility is **not** given for almost all settings.

6. Acknowledgements

This project received funding from the European Union's Horizon 2020 research and innovation program under grant agreements No. 690907 (IDENTITY).

References

[1] ISO/IEC 24745:2011 Information technology — Security techniques — Biometric information protection, 2011.

[2] ISO/IEC 30136:2018 Information technology — Performance testing of biometric template protection schemes, 2018.

[3] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (12):2037–2041, 2006.

[4] J. Bringer, H. Chabanne, and C. Morel. Shuffling is not sufficient: Security analysis of cancelable iriscodes based on a secret permutation. In *IEEE International Joint Conference on Biometrics*, pages 1–8. IEEE, 2014.

[5] J. Bringer, C. Morel, and C. Rathgeb. Security analysis and improvement of some biometric protected templates based on bloom filters. *Image and Vision Computing*, 58:239–253, 2017.

[6] C. H. Chan, X. Zou, N. Poh, and J. Kittler. Illumination invariant face recognition: a survey. In *Computer Vision: Concepts, Methodologies, Tools, and Applications*, pages 58–79. IGI Global, 2018.

[7] S. Chen, Y. Liu, X. Gao, and Z. Han. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In *Biometric Recognition*, pages 428–438, 2018.

[8] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2019.

[9] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou. Retinaface: Single-stage dense face localisation in the wild. *CoRR*, abs/1905.00641, 2019.

[10] C. Ding and D. Tao. A comprehensive survey on pose-invariant face recognition. *ACM Transactions on intelligent systems and technology (TIST)*, 7(3):37, 2016.

[11] P. Drozdowski, S. Garg, C. Rathgeb, M. Gomez-Barrcro, D. Chang, and C. Busch. Privacy-preserving indexing of iris-codes with cancelable bloom filter-based search structures. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 2360–2364. IEEE, 2018.

[12] Z. Erkin, M. Franz, J. Guajardo, S. Katzenbeisser, I. Lagendijk, and T. Toft. Privacy-preserving face recognition. In *International symposium on privacy enhancing technologies symposium*, pages 235–253. Springer, 2009.

[13] Y. Feng, P. C. Yuen, and A. K. Jain. A hybrid approach for face template protection. In *Biometric Technology for Human Identification V*, volume 6944, page 694408. International Society for Optics and Photonics, 2008.

[14] C. A. Glasbey and K. V. Mardia. A review of image-warping methods. *Journal of applied statistics*, 25(2):155–171, 1998.

[15] M. Gomez-Barrero, J. Galbally, C. Rathgeb, and C. Busch. General framework to evaluate unlinkability in biometric template protection systems. *IEEE Transactions on Information Forensics and Security*, 13(6):1406–1420, 2018.

[16] M. Gomez-Barrero, C. Rathgeb, J. Galbally, C. Busch, and J. Fierrez. Unlinkable and irreversible biometric template protection based on bloom filters. *Information Sciences*, 370:18–32, 2016.

[17] J. Hämmerle-Uhl, E. Pschernig, and A. Uhl. Cancelable iris biometrics using block re-mapping and image warping. In P. Samarati, M. Yung, F. Martinelli, and C. Ardagna, editors, *Proceedings of the 12th International Information Security Conference (ISC'09)*, volume 5735 of *LNCS*, pages 135–142. Springer Verlag, 2009.

[18] C.-K. Hsieh, S.-H. Lai, and Y.-C. Chen. Expression-invariant face recognition with constrained optical flow warping. *IEEE Transactions on Multimedia*, 11(4):600–610, 2009.

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

[20] R. Jafri and H. R. Arabnia. A survey of face recognition techniques. *Jips*, 5(2):41–68, 2009.

[21] S. Kirchgasser, C. Kauba, and A. Uhl. Cancellable biometrics for finger vein recognition - application in the feature domain. In S. M. R. V. Andreas Uhl, Christoph Busch, editor, *Handbook of Vascular Biometrics*, chapter 16, pages 507–525. Springer Nature Switzerland AG, Cham, Switzerland, 2019.

[22] P. Korshunov and T. Ebrahimi. Using warping for privacy protection in video surveillance. In *2013 18th International Conference on Digital Signal Processing (DSP)*, pages 1–6. IEEE, 2013.

[23] Y.-L. Lai, Z. Jin, A. B. J. Teoh, B.-M. Goi, W.-S. Yap, T.-Y. Chai, and C. Rathgeb. Cancellable iris template generation based on indexing-first-one hashing. *Pattern Recognition*, 64:105–117, 2017.

[24] Y. Martindez-Diaz, L. S. Luevano, H. Mendez-Vazquez, M. Nicolas-Diaz, L. Chang, and M. Gonzalez-Mendoza. Shufflefacenet: A lightweight face architecture for efficient and highly-accurate face recognition. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.

[25] Y. Martínez-Díaz, N. Hernandez, R. J. Biscay, L. Chang, H. Mendez-Vazquez, and L. E. Sucar. On fisher vector encoding of binary features for video face recognition. *Journal of Visual Communication and Image Representation*, 51:155–161, 2018.

[26] O. M. Parkhi, K. Simonyan, A. Vedaldi, and A. Zisserman. A compact and discriminative face track descriptor. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1693–1700, 2014.

[27] V. M. Patel, N. K. Ratha, and R. Chellappa. Cancelable biometrics: A review. *IEEE Signal Processing Magazine*, 32(5):54–65, 2015.

[28] E. Piciucco, E. Maiorana, C. Kauba, A. Uhl, and P. Campisi. Cancelable biometrics for finger vein recognition. In *Proceedings of the 1st Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE 2016)*, pages 1–6, Aalborg, Denmark, 2016.

[29] N. K. Ratha, J. H. Connell, and R. M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal*, 40(3):614–634, 2001.

[30] C. Rathgeb and A. Uhl. Two-factor authentication or how to potentially counterfeit experimental results in biometric systems. In *Proceedings of the International Conference on Image Analysis and Recognition (ICIAR'10)*, volume 6112 of *Springer LNCS*, pages 296–305, Povoa de Varzim, Portugal, June 2010.

[31] G. Wolberg. Image morphing: a survey. *The visual computer*, 14(8):360–372, 1998.

[32] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li. Face detection based on multi-block lbp representation. In *International conference on biometrics*, pages 11–18. Springer, 2007.