IET Research Journals

IET Biometrics

IET Journals
The Institution of Engineering and Technology

ISSN 1751-8644
doi: 0000000000
www.ietdl.org

Predicting the Vulnerability of Biometric Systems to Attacks based on Morphed Biometric Information

Marta Gomez-Barrero ✉, *Christian Rathgeb, Ulrich Scherhag, Christoph Busch*

[1] *da/sec - Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany*
✉ *E-mail: marta.gomez-barrero@h-da.de*

Abstract: Morphing techniques can be used to create artificial biometric samples or templates, which resemble the biometric information of two or more individuals in signal and feature domain. If morphed biometric samples or templates are infiltrated to a biometric recognition system, the subjects contributing to the morphed sample can be both successfully verified against a single enrolled template. Hence, the unique link between individuals and their biometric reference data is not warranted. This leads to serious security gaps in biometric applications, in particular the issuance and verification process of electronic travel documents. Recently, different biometric systems have been attacked using morphed biometric samples. However, so far a systematic approach to predict the vulnerability of the system to such attacks has not been proposed.
In this work, we present a framework to evaluate the vulnerability of biometric systems to attacks using morphed biometric information. Based on a biometric system's mated/non-mated score distributions and its decision threshold, a theoretical vulnerability assessment is proposed. In an experimental evaluation, the vulnerability of a face and an iris recognition system is quantified based on the presented framework. Obtained results are verified against real attacks based on morphed face images and morphed iris-based templates.

1 Introduction

Biometrics refers to the automated recognition of individuals based on their biological and behavioural characteristics [1]. Nowadays, biometric technologies represent an integral component of identity management and access control systems, providing a strong and permanent link between individuals and their identity. In past years, researchers have pointed out diverse potential vulnerabilities of biometric recognition systems. Proposed attacks, which aim at gaining unauthorized access to the system, can be coarsely categorized into presentation attacks and software-based attacks [2]. Presentation attacks refer to a presentation of an attack instrument (e.g. print outs or electronic displays [3]) to the biometric sensor with the goal of interfering with the operation of the biometric recognition system [4]. To launch software attacks, e.g., substitution attacks or overriding one of the inner modules of the system, an attacker requires knowledge about the interior architecture of the biometric system together with access to some of the system components.

Recently, attacks on face, fingerprint and iris recognition systems based on morphed biometric images and templates have been presented [5–9]. Morphed biometric information is an artificially generated sample or template, which blends the biometric information of two different data subjects into one. If such a morphed information is infiltrated into a biometric system at the time of enrolment, there is a high chance that the data subjects contributing to the morphed sample or template are successfully verified against it employing state-of-the-art recognition systems, i.e. the desired unique link between subject and the template is annulled. Fig. 1 shows the diagram of such an attack for face in the image domain, which is carried out in the following four steps: (1) the attacker finds an accomplice, whose biometric characteristic is similar enough to his own; (2) their characteristics are captured resulting in the biometric samples $\mathbf{M}_{ac}$ and $\mathbf{M}_{at}$, respectively; (3) a morphed sample $\mathbf{M}_{morph}$ is created from the original unaltered samples of the accomplice and the attacker, $\mathbf{M}_{ac}$ and $\mathbf{M}_{at}$; (4) the morphed sample $\mathbf{M}_{morph}$ is presented to the system, and its corresponding template $\mathbf{T}_{morph}$ is enrolled in the database. Later on, both the attacker and the accomplice can present their unaltered biometric characteristics to the biometric system, and the samples $\mathbf{M}'_{at}$ and $\mathbf{M}'_{ac}$ are captured, which will yield the corresponding templates $\mathbf{T}'_{at}$ and $\mathbf{T}'_{ac}$. The attack will be successful if both templates obtain similarity scores with respect to the enrolled morphed template higher than the decision threshold, δ:

$$s_{at} = S(\mathbf{T}'_{at}, \mathbf{T}_{morph}) > \delta \wedge s_{ac} = S(\mathbf{T}'_{ac}, \mathbf{T}_{morph}) > \delta \quad (1)$$

where S outputs the similarity score between two templates computed by the biometric system. Hence, an attack attempt is considered successful if $\min(s_{at}, s_{ac}) > \delta$.

Such attacks pose severe security threats to biometric systems, in particular to the issuance and verification process of electronic travel documents [5]: black-listed criminal offenders can use an authentic passport, complying with all physical document security features, to enter a country with the identity of an accomplice, when performing three basic steps: (1) find a rather lookalike accomplice, (2) morph passport face photos of both, possibly utilizing free software available on the internet, and (3) the accomplice applies for a passport. The passport manufacturer will issue an authentic passport equipped with the morphed biometric information and other identity attributes of the accomplice, which can be used to enter a country by both subjects. It should be noted, that the accomplice needs to cooperate with the attacker in order to apply for the valid passport containing his biographic data and the morphed biometric sample.

Different commercial face recognition systems have been found to be highly vulnerable to this type of attack [5]. Due to a high intra-class variability in human faces, face recognition systems are operated at false match rates (FMRs) as high as 0.1% to achieve acceptable false non-match rates (FNMRs). At such big FMRs, the chance that different subjects exhibit similar biometric features or even yield a biometric collision, becomes alarmingly high even for a rather small number of registered subjects. That is, it is expected to be straightforward for an attacker to find a suitable accomplice. Apart from relatively high FMRs, it is of particular interest what makes a biometric system potentially vulnerable to attacks based on morphed biometric information. Ideally, such an assessment should

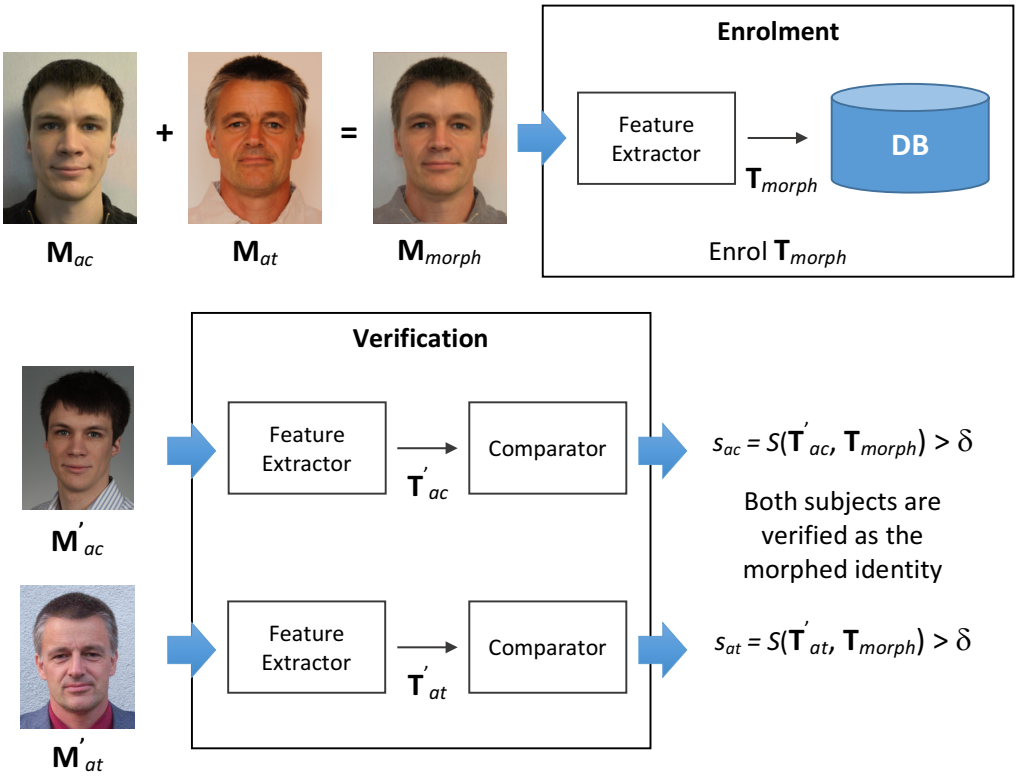


Fig. 1: An attack based on morphed facial images: at enrolment $\mathbf{M}_{at}$ and $\mathbf{M}_{ac}$ are morphed into $\mathbf{M}_{morph}$, and template $\mathbf{T}_{morph}$ is enrolled in the database. During verification, when either probe sample $\mathbf{M}'_{at}$ and $\mathbf{M}'_{ac}$ is presented to the system, their corresponding templates achieve similarity scores $s_{ac} = S\left(\mathbf{T}'_{ac}, \mathbf{T}_{morph}\right)$ (resp. at), higher than the decision threshold δ.

be based on biometric systems' key factors without the need of conducting specific attacks.

1.1 Contribution of Work

In this work we propose a theoretical framework for predicting the vulnerability of biometric systems to attacks based on morphed biometric information, which extends the framework introduced in [10]. It allows an assessment of the impact of said attacks for different operating points of biometric systems, i.e., decision thresholds δ. This evaluation is based on the relationship between mated and non-mated score distributions. Hence, it only requires the computations of the mated and non-mated scores of unaltered biometric samples, which is always necessary to fix the decision threshold of the system. In contrast to the framework presented in [10], in which the vulnerability assessment is based on the entire shape of score distributions, a pair-wise subject-specific analysis is conducted. By quantifying the necessary closeness of a pair of biometric samples, $\mathbf{M}_{at}$ and $\mathbf{M}_{ac}$, in order to launch a successful attack, appropriate values of δ can be chosen to achieve more robustness. Alternatively, modules designed to detect morphed biometric samples can be incorporated to the biometric system, where such detection algorithms are currently developed by different research teams [6, 8, 11, 12].

The soundness of the presented framework is tested by conducting different attacks based on morphed biometric information, thereby extending the work of [10]. Whereas facial information is morphed in image or sample domain, iris information is morphed in feature domain.

Therefore, the main contributions can be summarised as follows:

- Improved evaluation framework for the vulnerabilities of biometric systems to attacks based on morphed information. In particular, a pair-wise subject-specific analysis is added to the approach proposed in [10].
- Two different case studies are analysed with the proposed framework in the experimental section: (1) face sample level morphing and (2) iris feature level morphing.
- In addition, the results of the estimation framework are compared to the empirical evaluation carried out with the metrics described in [13].
- The experimental evaluation has been carried out on open source systems and publicly available databases, thus facilitating reproducibility.

1.2 Article Organisation

The rest of this work is organised as follows: Sect. 2 briefly summarizes related works with respect to attacks based on morphed biometric information. Subsequently, some key definitions are summarised in Sect. 3 and the proposed framework used to predict the vulnerability of a biometric system to such attacks is described in detail in Sect. 4. The metrics used to evaluate specific attacks are listed in Sect. 5 and experimental evaluations are presented in Sect. 6. Finally, conclusions are drawn in Sect. 7.

2 Related Works

Attacks based on morphed biometric samples were first introduced by Ferrara *et al.* [5]. Motivated by security gaps in the issuance process of electronic travel documents, the authors showed that commercial face recognition software tools are highly vulnerable to such attacks, i.e. different instances of images of either subject are successfully matched against the morphed image. In their experiments, decision thresholds yielding a false match rate (FMR) of 0.1% have been used, according to the guidelines provided by the European Agency for the Management of Operational Cooperation at the External Borders (FRONTEX) [14]. In a further study, the authors show that morphed face images are realistic enough to fool human examiners [15]. Scherhag *et al.* [6] showed that presentation attack detection schemes employing general purpose texture

descriptors used in conjunction with machine learning techniques are not capable of reliably detecting morphed face images. With respect to the above attack scenario, it is stressed that a detection of morphed face images becomes even more challenging if images are printed and scanned. Hildebrandt *et al*. [8] suggest to employ generic image forgery detection techniques, in particular multi-compression anomaly detection, to reliably detect morphed facial images. Kraetzer *et al*. [12] evaluate the feasibility of detecting facial morphs with keypoint descriptors and edge operators. The benefits of deep neural networks for detecting morphed images have been recently investigated in [16, 17]. Subsequently, Raghavendra *et al*. explore in [18] the effectiveness of a collaborative representation of micro-texture features extracted from the colour space for face morphing detection. Agarwal *et al*. show in [19] that a novel approach to morphing detection based on a Weighted Local Magnitude Pattern feature descriptor outperforms several existing approaches for a newly created database, which will be publicly available in the future. In [20], Makrushin *et al*. explore facial morphing detection based on the distribution of Benford features extracted from quantized DCT coefficients of JPEG-compressed morphs and bona fide images. Neubert [21] analyses a continuous image degradation approach for morphing detection. Finally, in [22], Wandzik *et al*. study the vulnerabilities of a CNN based face recognition system towards morphing attacks, depending on the contribution of each subject to the morphed sample.

Regarding other biometric characteristics, Ferrara *et al*. [7] also presented two different methods to morph fingerprints in image and feature domain. For a decision threshold correspoding to a FMR of 0.1%, it is shown that commercial fingerprint recognition systems are also highly vulnerable to such attacks. Since fingerprint enrolment is usually done live in the issuance process of electronic travel documents, the authors argue that manufactured fake fingertips may be presented. More recently, Rathgeb and Busch [9] presented a technique to create morphed iris-based templates. It is shown that even iris recognition systems, which operate at a FMR of 0.0001%, might be vulnerable to said attacks.

Gomez-Barrero *et al*. [10] proposed the first theoretical framework for measuring the vulnerability of biometric systems to these attacks. Evaluations are conducted for diverse biometric systems, where expected comparison scores of attacks based on morphed images or templates are directly derived from the mated and non-mated distributions of a face, fingerprint and iris recognition system. The authors identified key factors which have a major influence on a system's vulnerability to such attacks, e.g. the shape of mated and non-mated score distributions or the FMR the system is operated at. Since there is no standardised manner to evaluate the vulnerability of biometric systems to attacks based on morphed information, Scherhag *et al*. [13] introduced new metrics for vulnerability reporting (further details in Sect. 6), which strongly relate to the metrics defined in [4], and which will be employed in our experiments. In addition, the authors provide recommendations on the assessment of morphing techniques. It is emphasized that unrealistic assumptions with respect to the quality of morphed biometric samples might cloud the picture regarding the performance of detection algorithms. In summary it becomes clear that research on attacks based on morphed biometric samples is still in statu nascendi. Nonetheless, at the time of this writing we see an increasing interest in this topic and the results of ongoing activities of different research labs are expected to be presented across diverse platforms in the near future.

3 Definitions and Notations

To extend formality to the problem being addressed, some notations are introduced in this section. Throughout the article we will use the Harmonized Biometric Vocabulary (HBV) defined in the ISO/IEC 2382-37 [1]. Given that they are often used throughout the article, for the sake of clarity, we include here the next definitions:

- *Biometric characteristic*: "biological and behavioural characteristic of an individual from which distinguishing, repeatable biometric features can be extracted for the purpose of biometric recognition". For example, a fingerprint or an iris are two different biometric characteristics.
- *Biometric instance*: for some characteristics, an individual possesses several instances. For example, the right index fingerprint is a different instance from the left thumb, even if they serve to identify the same person.
- *Mated samples*: "paired biometric probe and biometric reference that are from the same biometric characteristic of the same biometric data subject". For example, two samples from the same iris.
- *Non-mated samples*: "paired biometric probe and biometric reference that are not from the same biometric instance". For example, two samples from different irises.
- *Bona fide presentation*: "interaction of the biometric capture subject and the biometric data capture subsystem in the fashion intended by the policy of the biometric system". In other words, bona fide is analogous to normal or routine, in contrast to the presentation of a synthetic artefact, such as a morphed photo.
- *Template*: "set of stored biometric features comparable directly to probe biometric features". For instance, the binary iris-code extracted from an iris image.
- *Similarity score*: "numerical value (or set of values) resulting from a comparison" which increases with similarity. That is, given a comparison function S, the similarity score s between two templates $\mathbf{T}_1$ and $\mathbf{T}_2$ is defined as: $s = S(\mathbf{T}_1, \mathbf{T}_2)$.

In general, depending on the bona fide samples compared, two different types of similarity scores are possible within a biometric system: those obtained from the comparison of mated samples, and those yielded by comparisons of non-mated samples. Let us accordingly define the corresponding types of score distributions, where $s = S(\mathbf{T}_1, \mathbf{T}_2)$ is the similarity score between two templates:

- *Mated* trial distribution: scores computed from templates extracted from different samples of a single biometric instance of the same subject. It represents the conditional probability of obtaining a score s knowing that two templates come from mated samples representing the same biometric instance, that is, $p\left(s|H_m \wedge H_{bf}\right)$, where

$$H_m = \{\text{both templates stem from mated samples}\}$$
$$H_{bf} = \{\text{both templates stem from bona fide samples}\}$$

- *Non-mated* trial distribution: scores yielded by templates generated from samples of different instances. It represents the conditional probability of obtaining a score s knowing that two templates come from non-mated samples (i.e. representing not the same instance), that is, $p\left(s|H_{nm} \wedge H_{bf}\right)$, where

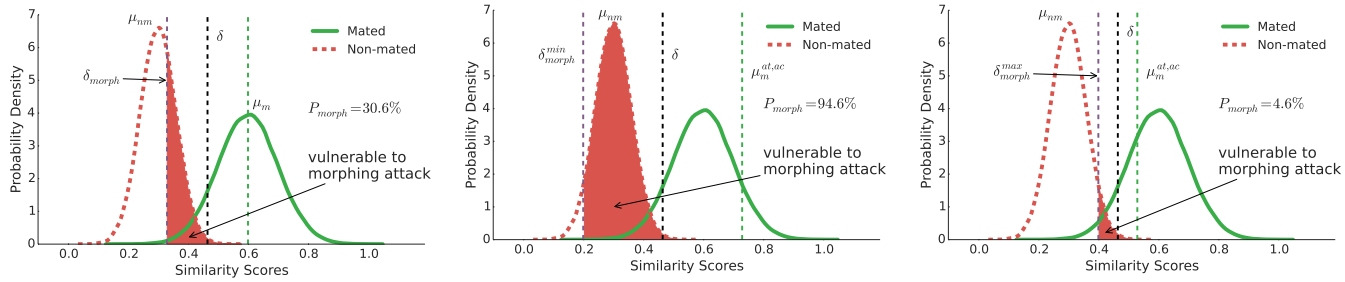
$$H_{nm} = \{\text{both templates stem from non-mated samples}\}$$

Two examples of the probability density functions of the mated and non-mated distributions are shown in Fig. 2, where the *Non-mated* samples distribution, $p\left(s|H_{nm} \wedge H_{bf}\right)$, is depicted in dashed red, and the *Mated* samples distribution, $p\left(s|H_m \wedge H_{bf}\right)$, in solid green, their corresponding mean values are denoted as μ_m and μ_{nm}, respectively, and the decision threshold δ is represented with a vertical black dashed line.

For the problem at hand, where bona fide templates can be compared to either morphed templates or templates extracted from morphed samples, a third score distribution is defined:

- *Morphed* trial distribution: the comparison of a template generated from morphed biometric information to another independent bona fide template of one contributing subject, that is, $p\left(s|H_{nm} \wedge H_{morph}\right)$, where

$$H_{morph} = \{\text{one template stems from morphed information}\}$$



(a) $\mu = \mu_m$ (b) $\mu = \mu_m^{at,ac}$ and δ_{morph}^{min} (c) $\mu = \mu_m^{at,ac}$ and δ_{morph}^{max}

Fig. 2: Example of *Mated* (solid green) and *Non-mated* (dashed red) trial distributions. In each case, the corresponding P_{morph} has been computed using a single μ (dashed green line), different for each subfigure: (a) the mean value of the mated score distribution, μ_m, and the mean value for the mated scores of two particular subjects, $\mu_m^{at,ac}$, which yield (b) the minimum morph threshold, δ_{morph}^{min}, and (c) the maximum morph threshold, δ_{morph}^{max}. In all cases P_{morph} corresponds to the shaded area below $p\left(s|H_{nm} \wedge H_{bf}\right)$.

4 Theory: General Framework

In order to assess the feasibility of carrying out attacks based on morphed information such as the ones described in Sect. 1 and Fig. 1, we have to answer the following question: what is the probability, denoted as P_{morph}, that both subjects contributing to the morphed image or template are positively matched with it? In other words, for a given biometric system we want to compute

$$P_{morph} = p\left(\min\left(s_{at}, s_{ac}\right) > \delta\right) \tag{2}$$

where δ denotes the decision threshold and s represents the similarity score output by the biometric system during the decision process. s is the only variable present in all equations in the remainder of the section, and will have a range determined by the biometric system evaluated. For simplicity, in the experimental section (Sect. 6), s is normalised to the range [0, 1].

In practice, one of the aforementioned scores can be higher than the other one (i.e., $s_{at} \neq s_{ac}$). However, since the relationship between s_{at} and s_{ac} depends on the nature of the morphing process (i.e., the weight of each subject on the final morphed sample), and the main goal of the theoretical framework is to evaluate automatic recognition systems (i.e., the morphed samples do not need to fool a human examiner but only a software tool), we assume that $s_{at} \approx s_{ac}$ (i.e., each subject has an equal weight of 50%) as in [10]. With that assumption, the success probability P_{morph} defined in Eq. 2 is maximised, thereby evaluating the worst-case scenario, or highest success chances of the attack. We thus refer to any of the scores as s_{at}.

The success of these attacks depends on where s_{at} lies with respect to the decision threshold δ: the attack will only be successful if the identity claim is accepted, that is, if $s_{at} > \delta$. Since s_{at} stems from a non-mated trial (i.e., the attacker or the accomplice against the morphed sample, which represents a third instance), it will belong to the *Non-mated* distribution. However, for similarity scores it is more probable that s_{at} lies on the right tail of the *Non-mated* distribution, between the mean values of both score distributions, μ_m and μ_{nm}. This is due to the the following reasons:

- either the reference template $\mathbf{T}_{morph}$ is extracted from $\mathbf{M}_{morph}$ (i.e., morphing at sample level), which is ultimately a combination of $\mathbf{M}_{at}$ and $\mathbf{M}_{ac}$,
- or the reference template $\mathbf{T}_{morph}$ has been created as a combination of $\mathbf{T}_{at}$ and $\mathbf{T}_{ac}$ (i.e., morphing at feature level).

In both cases, it is assumed that the morph was created to allow a positive verification of both subjects, and as a consequence $\mathbf{T}_{morph}$ lies between both bona fide templates $\mathbf{T}_{ac}$ and $\mathbf{T}_{at}$ in the feature space. Due to the assumption of $s_{at} \approx s_{ac}$ and also assuming that the comparator has a 'quasi linear' behaviour (e.g., based on a distance metric), it is expected to be close to the average of the *Mated* and *Non-mated* scores. Thus, as established in [10], if for a given accomplice whose characteristic yields a non-mated similarity score s_{nm} with respect to the attacker:

$$s_{nm} = S\left(\mathbf{T}'_{ac}, \mathbf{T}'_{at}\right) \tag{3}$$

the expected value of s_{at} (μ_{at}) can be estimated as:

$$\mu_{at} = E\left(s_{at}\right) = E\left(\frac{s_{nm} + s_m}{2}\right) = \frac{s_{nm} + \mu_m}{2} \tag{4}$$

where $s_m = S\left(\mathbf{T}'_{morph}, \mathbf{T}_{morph}\right)$ represents a mated score, and hence has an expected value of μ_m.

Such an approximation only takes into account the similarity of the contributing subjects, s_{nm}. However, the nature of the subjects is not reflected. In other words, the expected mated similarity score of the attacker in Eq. 4 is modelled as the mean mated score of all the enrolled subjects, μ_m. Nonetheless, the morph of two subjects, whose samples always obtain very high mated similarity scores, is also expected to achieve high morphed similarity scores, and vice versa. The reason behind such an assumption lies on the nature of the morphed template, which is expected to lie between both bona fide templates in the feature space. Hence, a more precise estimation can be obtained if μ_m is substituted by the mean value of the mated scores stemming from the accomplice and the attacker:

$$\mu_m^{at,ac} = \overline{\mathcal{S}^{at,ac}} \tag{5}$$

where $\mathcal{S}^{at,ac} = \left\{s_m^{1,at}, \ldots, s_m^{I^{at},at}\right\} \cup \left\{s_m^{1,ac}, \ldots, s_m^{I^{ac},ac}\right\}$, and I^{at} (resp. I^{ac}) indicates the number of mated scores of the attacker (resp. accomplice).

Therefore, the morph score can be estimated as:

$$\mu_{at} = E\left(s_{at}\right) = E\left(\frac{s_{nm} + s_m}{2}\right) = \frac{s_{nm} + \mu_m^{at,ac}}{2} \tag{6}$$

Now, the probability of success of the morphing attack, as defined in Eq. 2, ultimately depends on the chances of obtaining an accomplice for which μ_{at} lies above the decision threshold δ. Which in turn depends on the score yielded by the accomplice with respect to the attacker, s_{nm}, and the average mated score of both subjects, $\mu_m^{at,ac}$:

$$\begin{aligned} P_{morph} &= P\left(\mu_{at} > \delta\right) = P\left(\frac{s_{nm} + \mu_m^{at,ac}}{2} > \delta\right) \\ &= P\left(s_{nm} > 2\delta - \mu_m^{at,ac}\right) \end{aligned} \tag{7}$$

Denoting

$$\delta_{morph}^{at,ac} = 2\delta - \mu_m^{at,ac} \tag{8}$$

we can finally compute P_{morph} as follows:

$$\begin{aligned} P_{morph} &= \sum_{at} \sum_{ac \neq at} \left\{ \frac{s_{nm} + \mu_m^{at,ac}}{2} > \delta \right\} \\ &= \sum_{at} \sum_{ac \neq at} \left\{ s_{nm} > 2\delta - \mu_m^{at,ac} \right\} \\ &= \sum_{at} \sum_{ac \neq at} \left\{ s_{nm} > \delta_{morph}^{at,ac} \right\} \end{aligned} \tag{9}$$

It should be noted that a different threshold $\delta_{morph}^{at,ac}$ is computed for each pair of subjects. In the experimental section, the minimum δ_{morph}^{min} and maximum δ_{morph}^{max} thresholds will be reported.

Finally, Fig. 2a shows the impact of utilising a single value of μ for the final estimation of P_{morph}. In all cases, a single value of μ (green dashed line) has been utilised in order to facilitate the visualization. The corresponding δ_{morph} is represented with a purple vertical dashed line, and the area for which $s_{nm} > \delta_{morph}$ (i.e., successful attack, representing P_{morph}) is shaded in red. As it may be observed, if we assume that both constituting subjects will achieve a mated score μ_m (Eq. 4, Fig. 2a), as suggested in [10], the probability of success is $P_{morph} = 30.6\%$. However, if both subjects yield mated scores higher than μ_m (Fig. 2b), the probability of success rises to $P_{morph} = 94.6\%$. The opposite occurs when $\mu_m^{at,ac} < \mu_m$ (Fig. 2c, $P_{morph} = 4.6\%$).

In order to facilitate the use of the present framework and allow reproducibility of the article, a Python implementation of P_{morph} will be made available through the da/sec website and the da/sec Github account.

5 Practice: Evaluation Metrics

Regarding experimental evaluation metrics for specific morphing methods, Scherhag *et al.* proposed in [13] the Mated Morph Presentation Match Rate (MMPMR). This metric is an adaptation of the general Impostor Attack Presentation Match Rate (IAPMR) introduced in ISO/IEC 30107-3 [23], which is defined as the proportion of attack presentations using the same presentation attack instrument species in which the target reference is matched. The specificities of the attacks carried out with morphed information, which differ from presentation attacks, are captured as follows:

- If multiple bona fide templates of one subject are compared to one morphed template, such comparisons can be understood as multiple authentication attempts per subject. The subject is thus successfully verified as long as one attempt is above the threshold of the biometric system. In other words, only the *maximum* of such scores is considered:

$$\max_{i=1,\ldots,I_m^{at}} s_m^{at,i} \tag{10}$$

where $s_m^{at,i} = S\left(\mathbf{T}_{at}^i, \mathbf{T}_{morph}^m\right)$ is the similarity score between the template extracted from the i-th sample of the attacker ($\mathbf{T}_{at}^i$), and the m-th morphed template $\mathbf{T}_{morph}^m$. In total, I_m^{at} trials are carried out, which in practice would be specified in the security policy of the operational system.

- In addition, these attacks can be considered successful if *all* contributing subjects are positively verified against the morphed template. As a consequence, only the *minimum* similarity score of all morph trials against one morphed sample is considered:

$$\min \left[\max_{i=1,\ldots,I_m^{at}} s_m^{at,i}, \max_{i=1,\ldots,I_m^{ac}} s_m^{ac,i} \right] \tag{11}$$

Fig. 3: Sample images of two subjects (top and bottom row) of the AR face database [24].

Therefore, the MinMax-MMPMR metric can be defined as the average percentage of successful attacks:

$$\text{MinMax-MMPMR} = \frac{1}{M} \cdot \sum_{m=1}^{M} \left\{ \left(\min \left[\max_{i=1,\ldots,I_m^{at}} s_m^{at,i}, \max_{i=1,\ldots,I_m^{ac}} s_m^{ac,i} \right] \right) > \delta \right\} \tag{12}$$

It should be noted that the MMPMR depends on the decision threshold δ of the biometric system. Therefore, it is proposed in [13] to additionally report the Relative Morph Match Rate (RMMR), which represents the relationship between the percentage of correctly matched subjects ($1 - \text{FNMR}$) and the percentage of wrongly matched morph trials (MMPMR):

$$\text{RMMR}(\delta) = 1 + (\text{MMPMR}(\delta) - (1 - \text{FNMR}(\delta))) \tag{13}$$

An implementation of these metrics can be found at the da/sec Github account*.

6 Experimental Evaluation

6.1 Protocol

Two open-source systems will be analysed:

- **Face recognition**: the OpenFace [26] algorithm has been used. Face images are pre-processed to align them based on facial landmarks extracted with dlib [27]. To that end, the corners of the eyes and the centre of the nose are used as reference. Then, resized and cropped images of 96×96 pixels are fed to the default pretrained Deep Neural Network (DNN), to obtain a 128 dimensional face representation. Finally, templates are classified using a Support Vector Machine (SVM). For more details on the recognition method, such as the DNN model, the reader is referred to [26].
- **Iris recognition**: first, the iris is detected in the image, and transformed into a normalized rectangular texture of 512×64 pixels. During feature extraction, the normalized enhanced textures are divided into stripes and adjacent rows averaged in order to obtain 10 one-dimensional signals. Then, a quadratic spline wavelet transform is applied to obtain a final iris-codes of size 512×10 bits, as

*https://github.com/dasec/mvr

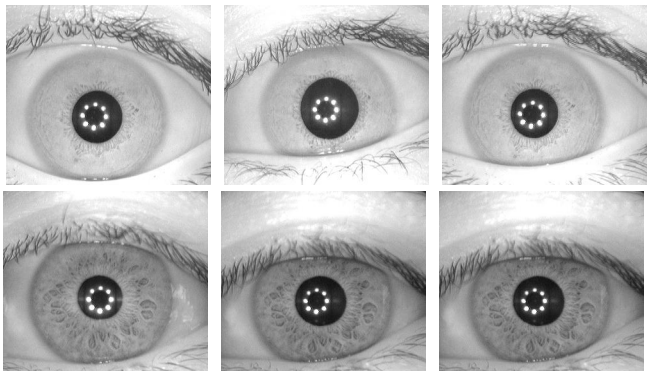


Fig. 4: Sample images of two eyes (top and bottom row) of the CASIAv4-Interval iris database [25].

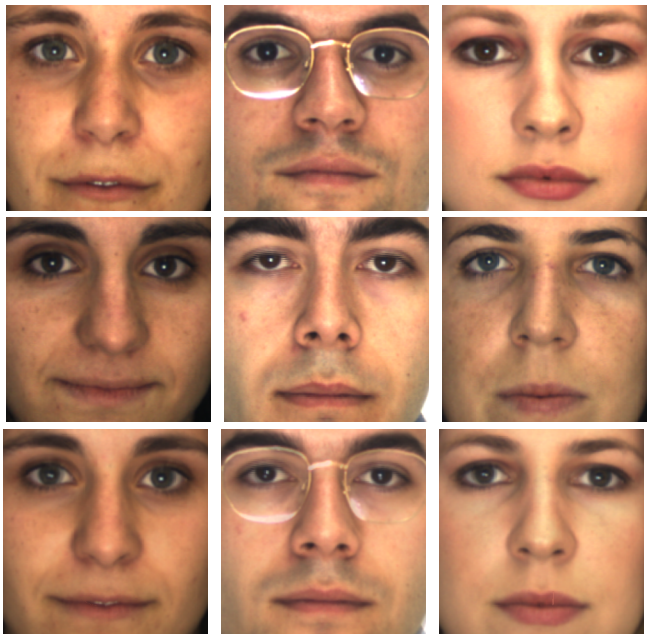


Fig. 5: Sample pre-processed face morphs (bottom) with the corresponding consituting images (top and centre).

proposed by Ma *et al.* [28]. The employed implementations of the algorithms are available in [29], and for further details the reader is referred to [30].

Two publicly available databases are used in the experiments, being in both cases all possible mated and non-mated trials computed:

- **Face - ARface database** [24]: from of the 136 subjects, the neutral, smiling und anger expresion images are considered (6 per subject). Morphs are generated from pairs of neutral images (first sample). Some samples are included in Fig. 3. The decision threshold selected corresponds to FMR = 0.1%, as it is recommended by FRONTEX [14].
- **Iris - CASIAv4-Interval iris database** [25]: we have taken into account images of all 198 left eyes, for which sample images are depicted in Fig. 4. We consider a Hamming distance of $\delta = 0.32$ as decision criterion which was recommended in [31], and which is expected to correspond to a FMR of 0.0001% (1 in a million).

Finally, two morphing techniques carried at different levels are evaluated:

- **Face morphing - image domain**: morphs are generated in a two step approach. First, facial landmarks are detected with dlib [27]. Those landmarks are subsequently used to morph both images using Delaunay triangulation [32] and Alpha Blending [33]. These morphed images can be presented at enrolment, and hence the corresponding morphed template will be stored as reference in the database. 8,853 morphed images are created, from which three examples are shown in Fig. 5.
- **Iris morphing - feature domain**: as proposed in [9], given a pair of iris-codes, $\mathbf{T}_{ac}$ and $\mathbf{T}_{at}$, and their corresponding noise masks, a morphed iris-code $\mathbf{T}_{morph}$ and a noise mask are created. To that end, entire rows are interleaved chosen from $\mathbf{T}_{ac}$ or $\mathbf{T}_{at}$, and assigned to $\mathbf{T}_{morph}$. In that selection, it is ensured that the same number of rows are chosen from both contributing iris-codes. Such morphed iris-codes can be infiltrated in the database at any time, or in the communication channel between the feature extractor and the database at enrolment, which are two of the vulnerable points of biometric systems [2]. A total number of 34,410 morphed iris-codes are created from pairs of the first image of each subject, from which an example is shown in Fig. 6.

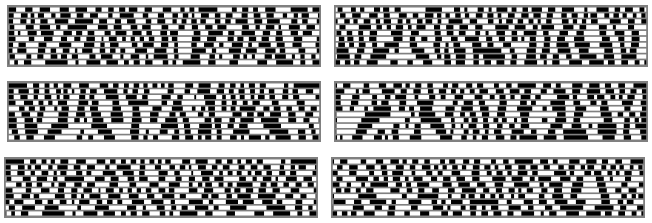


Fig. 6: Sample iris-code morph (bottom) with the corresponding consituting iris-codes (top and centre).

6.2 Results

In order to properly analyse the success chances of the attacks carried out with morphed samples, all morph scores are shown in Fig. 7, where they are represented with purple crosses, for face (top) and iris (bottom). The corresponding decision threshold, δ, is depicted with a dashed black line. In addition, we have included the theoretical scores (red stars), computed from the non-mated trials and the average mated score of the constituting samples (see Eq. 6). The percentage of scores which lay over the decision threshold (i.e., which grant a positive match) is included in the legend.

First, all possible combinations of similarity scores of the attacker and the accomplice with respect to the morphed sample are depicted as a set of points on the left figures (Figs. 7a and 7c, comprising 191,281 score pairs for face and 128,978 for iris). As it may be observed, in the practical evaluation, the percentage of positive matches is 0.42% for iris and 2.45% for face. However, on the right figures (Figs. 7b and 7d) only the relevant pairs of scores, as defined in Eq. 11, are depicted (8,853 pairs for face and 34,410 for iris, one pair for each morphed sample). In this case, a morphed image is successful if *any* verification attempt results in a match, and therefore the success chances increase to 1.95% and 9.58%, respectively. These differences highlight the importance of considering only the scores defined as in Eq. 11 in order to better estimate the success chances of the attack.

In addition to the aforementioned fact, we can also observe in the scatter plots that morphs which achieve high (resp. low) morph scores for one of the constituting subjects, are more likely to achieve it for both subjects. This is reflected on a smaller deviation from the bisector (i.e., $s_{at} = s_{ac}$) for extreme similarity scores rather than for medium similarity scores.

Keeping those reflections in mind, *Mated* (solid green), *Non-mated* (dashed red) and *Morphed* (solid purple) trial distributions for both systems are depicted in Fig. 8. Only the relevant morphed scores (see Eq. 11) are considered in this plot. As before, the decision threshold δ is depicted with a black dashed line, and the minimum δ^{min}_{morph} and maximum δ^{max}_{morph} morph threshold are plotted in purple. The value of P_{morph} and MinMax-MMPMR, which corresponds to the shaded area below $p\left(s|H_m \wedge H_{morph}\right)$,

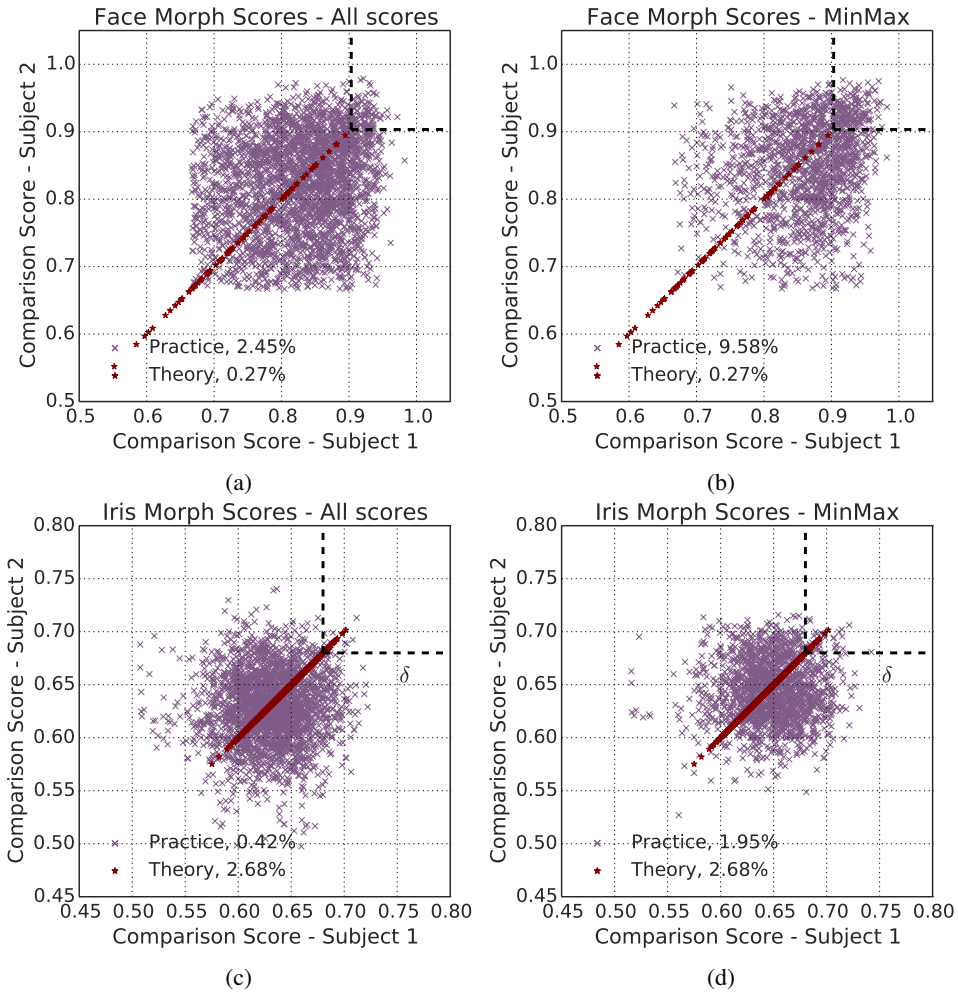


Fig. 7: Scatter plots of the morphed scores of a morph template with respect to both constituting subjects (purple 'x') and the theoretical scores s_{at} (red stars). In addition, the decision threshold δ is shown with a dashed black line. While on the left (a,c) all practical scores are depicted, only the relevant scores (see Eq. 11) are shown on the right (b,d). The percentages in the legend show the number of scores above the decision threshold δ.

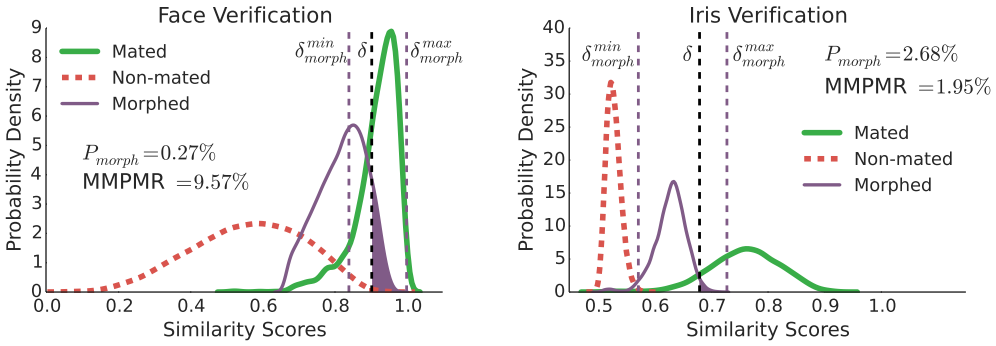


Fig. 8: *Mated* (solid green), *Non-mated* (dashed red) and *Morph* (solid purple) trial distributions for the systems based on face (left) and iris (right). The latter only consideres the relevant scores (Eq. 11). In all cases, the decision threshold δ is depicted with a black dashed line, and the minimum δ^{min}_{morph} and maximum δ^{max}_{morph} morphing threshold are plotted in purple. The value of P_{morph} and MinMax-MMPMR, which corresponds to the shaded area below $p\left(s|H_m \wedge H_{morph}\right)$, are also included.

are also included. The results of both the theoretical and the practical evaluations are summarised in Table 1, which also shows all the distributions statistics in terms of mean and standard deviation, as well as the intermediate values to compute P_{morph}.

As it may be observed in this last set of figures and on the table, since only the relevant scores are considered, the MinMax-MMPMR coincides with the practical evaluation carried out in Figs. 7b and 7d. Similarly, the reported P_{morph} is exactly the percentage of successful attacks reported in Fig. 7. For the iris system, the estimation is considerably closer to MinMax-MMPMR than that proposed in [10] ($\bar{P}_{morph} = 0.003\%$ vs $P_{morph} = 2.68\%$), due to the use of a pairwise subject specific mean value $\mu_m^{at,ac}$ (see Eq. 5). This is however not the case for face, where the estimation is very similar to $\bar{P}_{morph}$ due to the sharpness of the mated score distribution with respect to

Table 1 Evaluation of the distributions depicted in Fig. 8, including the corresponding mean (μ) and standard deviation (σ), the operating point analysed in terms of FMR and FNMR, with the corresponding decision threshold δ. For the theoretical analysis, the morphing δ^{min}_{morph} and δ^{max}_{morph}(Eq. 8) thresholds and the probability of success of a morphing attack P_{morph} (Eq. 9) are shown. For comparison, $\bar{P}_{morph}$ as presented in [10] is also included. Finally, for the practical analysis, the MinMax-MMPMR (Eq. 12) and RMNMR (Eq. 13) are depicted.

	EER	μ_m	σ_m	μ_{nm}	σ_{nm}	FMR	FNMR	δ	δ^{min}_{morph}	δ^{max}_{morph}	P_{morph}	$\bar{P}_{morph}$ [10]	MMPMR	RMMR
Face	5.73%	0.92	0.06	0.56	0.16	0.1%	30.53%	0.90	0.84	1.00	0.27%	0.26%	9.58%	40.11%
Iris	0.71%	0.76	0.06	0.53	0.01	0.0001%	9.36%	0.68	0.57	0.73	2.68%	0.003%	1.95%	11.31%

Table 2 Evaluation of lookalike morphs, including the theoretical estimation of the probability of success of a morphing attack P_{morph} (Eq. 9) and the empirical MinMax-MMPMR (Eq. 12).

		P_{morph}	MMPMR
Face	All morphs	0.27% (0.003 · MMPMR)	9.58%
	50% best morphs	0.84% (0.005 · MMPMR)	15.25%
Iris	All morphs	2.68% (1.32 · MMPMR)	1.95%
	50% best morphs	2.69% (1.05 · MMPMR)	2.55%

the non-mated distribution ($\sigma_m = 0.06$ *vs.* $\sigma_{nm} = 0.16$). In addition, whereas P_{morph} yields a good estimate for MinMax-MMPMR (1.95%) for the iris morphs, it is further from the MinMax-MMPMR of the face morphs (9.58%). This is due to the morphing process carried out in each case. For iris, half of the features are selected from each subject, as it is assumed in Eq. 6. However, a blending process is carried out on the facial images, from which the features are subsequently extracted. The quantization done by the feature extractor (i.e., from the input image only 128 values are extracted) leads to a considerable reduction in the feature space (i.e, compression), which in turn moves the morphed template closer to both subjects in the feature space. The theoretical framework, as a consequence, underestimates the success chances of the attack.

Regarding the empirical evaluation, the RMMR is also included in Table 1. As it may be observed, the absolute difference between the corresponding MinMax-MMPMR is increased for the RMMR: it decreased from 40.11% for face to 11.31% for iris. This is due to the difference in the FNMR (see Eq. 13), which is 9.36% for iris (i.e., roughly one out of ten mated trials will be wrongly rejected) in contrast to the 30.53% for face (i.e., roughly one out of three mated trials will be wrongly rejected). This metric, therefore, reflects the balance between the vulnerability of the system to attacks carried out with morphed samples and the usability in terms of false non-matches.

To conclude the section, it should be noted that when the success probabilities of the theoretical (P_{morph}) and the practical (MinMax-MMPMR) evaluations are analysed for the iris system, we observe that $P_{morph} >$ MinMax-MMPMR. This is due to the theoretical assumption that $s_{at} = s_{ac}$, which is not fulfilled in practice. However, we may observe that for high similarity scores, the real morphed mated scores are close to the bisector of the corresponding scatter plot, thereby representing a smaller deviation from the theoretical assumption. Since in a realistic scenario, only similar subjects would be morphed (e.g., with the same gender or skin color), we may conclude that, if only such more realistic morphs are taken into account, P_{morph} would be closer to MinMax-MMPMR. To study this fact, a final analysis of the behaviour of both the MinMax-MMPMR and the proposed metric P_{morph} has been included in Table 2. Both empirical and estimated values have been computed for all morphs analysed in Table 1 and for the best 50% morphs in terms of how similar the accomplice and the attacker characteristics are, in terms of their similarity score. Those morphs will be referred to as lookalike morphs. As it could be expected, if we restrict the attacks to those lookalike morphs, the MMPMR rises from 9.58% to 15.35% for face, and from 1.95% to 2.55% for iris. Similarly, in line with the discussion above, the estimation obtained from P_{morph} is twice as accurate for face. And for iris, the estimation deviates only by 5% relative from the empirical MinMax-MMPMR.

7 Conclusions

In this work, we have analysed, from both a theoretical and a practical perspective, the vulnerabilities of two different state-of-the-art and freely available biometric systems to attacks carried out with morphed templates (i.e., morphing in the feature domain) and images (i.e., morphing in the signal domain). To that end, we have proposed a framework, which builds upon a prior work [10], in order to estimate the probability of all constituent subjects being positively matched to the morphed reference stored in the database. The theoretical framework only needs access to mated and non-mated samples similarity scores in order to compute such probability, and can be eventually applied to any database as long as high quality morphs are generated, as recommended in [13]. In addition, it offers a more accurate estimation than the initial approximation in [10] for feature level morphing, since it takes into account the nature of the constituting subjects. A similar behaviour is expected from signal level morphing where no features are lost by the morphing process or no additional noise is introduced: for instance, the fingerprint morphing proposed in [7], where half of the minutiae set belongs to each subject. Further efforts are however required for morphing approaches such as the facial one analysed, where feature extraction processes including a severe quantization of the input sample should be considered in the evaluation framework.

For clarity, the framework has been described for the case where only two subjects are morphed into a single sample. However, it could be easily extended to a morph on n subjects by introducing two modifications in Eq. 6: (1) the mean value of the mated scores of all constituting subjects should be considered (i.e., $\mu_m^{1,\dots,n}$ instead of $\mu_m^{at,ac}$), and (2) the mean of the distances between the constituting subjects, $\{s_{1,2}, s_{1,3}, \dots, s_{1,n}, \dots, s_{n-1,n}\}$, should replace s_{nm}. We will analyse this extension of the framework in future works.

The theoretical estimation is compared with two specific morphing techniques in the feature and the image domain, respectively. First, the morphed mated score distributions are analysed, proving the importance of modelling the attack scenario correctly. In other words, only the relevant scores should be considered, which take into account two facts: (1) all constituting subjects need to be matched to the morphed reference and (2) at least one verification attempt must be successful. For this practical analysis, we have used the evaluation metrics proposed in [13], which are aligned with the ISO/IEC IS 30107-3 on Biometric presentation attack detection and will hence allow a fairer comparison with future research works.

Finally, with the aim to make the article reproducible, an implementation of the metrics will be made public through the da/sec website and the da/sec Github account.

Acknowledgment

This work was supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within the Center for Research in Security and Privacy (CRISP, `www.crisp-da.de`).

8 References

1 ISO/IEC TC JTC1 SC37 Biometrics, *ISO/IEC 2382-37:2017 IT – Vocabulary – Part 37: Biometrics*, ISO and IEC, 2017.

2 N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.
3 S. Marcel, M. Nixon, and S. Z. Li, *Handbook of Biometric Anti-Spoofing*. Springer-Verlag New York, Inc., 2014.
4 ISO/IEC TC JTC1 SC37 Biometrics, *ISO/IEC IS 30107-1. Information Technology – Biometrics presentation attack detection – Part 1: Framework*, International Organization for Standardization, Mar. 2016.
5 M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *Proc. Int. Joint Conf. on Biometrics (IJCB)*, 2014, pp. 1–7.
6 U. Scherhag, R. Raghavendra, K. B. Raja, M. Gomez-Barrero, C. Rathgeb, and C. Busch, "On the vulnerability of face recognition systems towards morphed face attacks," in *Proc. Int. Workshop on Biometrics and Forensics (IWBF)*, 2017, pp. 1–6.
7 M. Ferrara, R. Cappelli, and D. Maltoni, "On the feasibility of creating double-identity fingerprints," *IEEE Trans. on Information Forensics and Security*, vol. 12, no. 4, pp. 892–900, 2017.
8 M. Hildebrandt, T. Neubert, A. Makrushin, and J. Dittmann, "Benchmarking face morphing forgery detection: Application of stirtrace for impact simulation of different processing steps," in *Proc. Int. Workshop on Biometrics and Forensics (IWBF)*, 2017, pp. 1–6.
9 C. Rathgeb and C. Busch, "On the feasibility of creating morphed iris-codes," in *Proc. Int. Joint Conf. on Biometrics (IJCB)*, 2017, pp. 1–6.
10 M. Gomez-Barrero, C. Rathgeb, U. Scherhag, and C. Busch, "Is your biometric system robust to morphing attacks?" in *Proc. Int. Workshop on Biometrics and Forensics (IWBF)*, 2017, pp. 1–6.
11 R. Raghavendra, K. B. Raja, and C. Busch, "Detecting morphed face images," in *Proc. Int. Conf. on Biometrics: Theory, Applications, and Systems (BTAS)*, 2016.
12 C. Kraetzer, A. Makrushin, T. Neubert, M. Hildebrandt, and J. Dittmann, "Modeling attacks on photo-ID documents and applying media forensics for the detection of facial morphing," in *Proc. Workshop on Information Hiding and Multimedia Security (IH& MMSec)*, 2017, pp. 21–32.
13 U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. Veldhuis, L. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, R. Breithaupt, R. Raghavendra, and C. Busch, "Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting," in *Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*, 2017, pp. 1–12.
14 "FRONTEX – Research and Development Unit: Best practice technical guidelines for automated border control (ABC) systems," 2012, version 2.0.
15 M. Ferrara, A. Franco, and D. Maltoni, "On the effects of image alterations on face recognition accuracy," in *Face Recognition Across the Imaging Spectrum*, T. Bourlai, Ed. Springer International Publishing, 2016, pp. 195–222.
16 R. Raghavendra, K. Raja, S. Venkatesh, and C. Busch, "Transferable deep-cnn features for detecting digital and print-scanned morphed face images," in *2017 IEEE Conf. on Computer Vision and Pattern Recognition Workshop (CVPRW)*, July 2017.
17 C. Seibold, W. Samek, A. Hilsmann, and P. Eisert, "Detection of face morphing attacks by deep learning," in *Proc. Int. Workshop on Digital Forensics and Watermarking (IWDW)*, 2017, pp. 107–120.
18 R. Raghavendra, K. Raja, S. Venkatesh, and C. Busch, "Face morphing versus face averaging: Vulnerability and detection," in *Proc. Int. Joint Conf. on Biometrics (IJCB)*, 2017.
19 A. Agarwal, R. Singh, M. Vatsa, and A. Noore, "SWAPPED! digital face presentation attack detection via weighted localmagnitude pattern," in *Proc. Int. Joint Conf. on Biometrics (IJCB)*, 2017.
20 A. Makrushin, T. Neubert, and J. Dittmann, "Automatic generation and detection of visually faultless facial morphs," in *Proc. Int. Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*, 2017, pp. 39–50.
21 T. Neubert, "Face morphing detection: An approach based on image degradation analysis," in *Proc. Int. Workshop on Digital Forensics and Watermarking (IWDW)*, 2017, pp. 93–106.
22 L. Wandzik, R. V. Garcia, G. Kaeding, and X. Chen, "CNNs under attack: On the vulnerability of deep neural neworks based fce recognition to image morphing," in *Proc. Int. Workshop on Digital Forensics and Watermarking (IWDW)*, 2017, pp. 121–135.
23 ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC FDIS 30107-3:2017, IT – Biometric presentation attack detection – Part 3: Testing and Reporting*.
24 A. Martinez, "The AR face database," CVC Tech. Report, Tech. Rep., 1998.
25 Chinese Academy of Sciences' Institute of Automation, "CASIA Iris Image Database V4.0 – Interval," http://biometrics.idealtest.org, 2010.
26 B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications," CMU School of Computer Science, Tech. Rep., 2016.
27 D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, 2009.
28 L. Ma, T. Tan, Y. Wang, and D. Zhang, "Efficient iris recognition by characterizing key local variations," *IEEE Trans. on Image Processing*, vol. 13, no. 6, pp. 739–750, 2004.
29 "USIT – University of Salzburg iris toolkit," http://www.wavelab.at/sources/Rathgeb16a, version 2.0.x.
30 C. Rathgeb, A. Uhl, and P. Wild, *Iris Recognition: From Segmentation to Template Security*, ser. Advances in Information Security. Springer Verlag, 2013, vol. 59.
31 J. Daugman, "Probing the uniqueness and randomness of iriscodes: Results from 200 billion iris pair comparisons," *Proc. of the IEEE*, vol. 94, no. 11, pp. 1927–1935, 2006.
32 B. Delaunay, "Sur la sphère vide. a la mémoire de george voronoi," *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na*, vol. 6, pp. 793–800, 1934.
33 T. Porter and T. Duff, "Compositing digital images," in *Proc. ACM Siggraph Computer Graphics*, vol. 18, no. 3, 1984, pp. 253–259.